

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RÉGRESSION ROBUSTE BAYÉSIENNE À L'AIDE DE
DISTRIBUTIONS À AILES RELEVÉES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

IAN SCHILLER

MAI 2008

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens d'abord à remercier mon directeur de mémoire Alain Desgagné de m'avoir offert la chance de travailler sur ce passionnant projet de maîtrise (meilleure chance la prochaine fois GARCH!). En particulier, je tiens à le remercier pour le soutien qu'il m'a fourni tant sur le plan scientifique que professionnel. Ses connaissances et ses talents de pédagogue ont certainement facilité ma compréhension du sujet et l'écriture de ce mémoire. Je le remercie aussi pour sa patience à l'égard du temps passé à lire et relire les nombreux brouillons que je lui ai soumis tout au long de ce projet. Je le remercie aussi pour sa grande disponibilité et son appui dans les moments les plus difficiles. Enfin, je ne pourrais passer sous silence son soutien financier qui m'a permis de m'investir à fond dans ce merveilleux projet.

Je remercie mes professeurs Glenn Shorrock, Serge Alalouf et tout particulièrement Sorana Froda pour leur formation universitaire et pour leur grande disponibilité tout au long de mes études tant au baccalauréat qu'à la maîtrise. Je remercie aussi mes collègues de classe Gabrielle Boucher et Vincent Benoliel pour les bons moments, mais aussi pour les moments de folies. En particulier, je les remercie tous les deux pour avoir bien voulu partager leurs notes de cours au moment de la naissance de ma fille. Ce fut un geste grandement apprécié.

Je désire remercier mes parents Manon et Normand ainsi que ma soeur Valérie pour leur amour inconditionnel et pour leur soutien indéfectible dans tous les projets que j'entreprends. Je les remercie pour toutes les belles valeurs qu'ils m'ont léguées et pour leurs encouragements soutenus. Je remercie Éric "Titan" Lachance pour sa précieuse amitié et son dévouement total. Je remercie aussi la grande famille : Papi et Mami Bélisle, Papi et Mami Schiller (qui me manquent beaucoup), ma tante Lou-lou ainsi que mes oncles Michel, Serge et Ron. Je remercie aussi ma belle-famille Suzanne et Gilles

pour leurs encouragements.

Enfin, un merci tout spécial à ma conjointe Marie-Hélène, qui depuis bientôt dix ans vit au rythme de mes sautes d'humeur liées à mes bons et moins bons moments académiques. Je la remercie pour sa patience qui a été mise à rude épreuve à maintes reprises alors que, bien des samedis soirs, j'ai dû lui préférer mon crayon et mes cahiers de notes. Je la remercie pour ses encouragements et pour son soutien combien important, sans lesquels ce mémoire n'aurait jamais vu le jour. Je la remercie de m'avoir donné une merveilleuse petite fille qui transpire la joie de vivre. En ayant accepté de mettre en veille sa propre formation académique pour me permettre de mener la mienne à terme et afin de s'occuper de notre fille, je lui serai toujours redevable. Je dois admettre que bien des circonstances ont fait en sorte que je me suis lancé dans cette merveilleuse expérience qu'est la maîtrise. Cependant, rien de tout cela ne se serait produit si Marie ne s'était pas montrée insistante. Pour toutes ces raisons et bien plus, je te dédie ce mémoire.

En terminant, j'aimerais simplement dire à ma petite fille Chloé : maintenant, ton Papa s'en vient jouer avec toi petite Chouette!!

TABLE DES MATIÈRES

LISTE DES TABLEAUX	vi
LISTE DES FIGURES	vii
RÉSUMÉ	xiii
INTRODUCTION	1
CHAPITRE I	
MÉTHODES DE RÉGRESSION ROBUSTE	4
1.1 Contexte théorique	4
1.2 Estimateur des moindres carrés	6
1.3 Estimateurs-M	14
1.3.1 Méthode itérative	16
1.3.2 Exemples d'estimateurs-M	17
1.4 Droite résistante de Tukey	47
1.4.1 Algorithme	50
CHAPITRE II	
MÉTHODE D'ESTIMATION BAYÉSIENNE AVEC PARAMÈTRE D'ÉCHELLE CONNU	63
2.1 Contexte théorique	63
2.2 Contexte bayésien	63
2.3 Le problème de position	65
2.4 Aperçu des avancées en robustesse bayésienne	67
2.5 Notions sur les densités à ailes relevées	68
2.5.1 Conditions d'épaisseur et de régularité pour les ailes d'une densité	68
2.5.2 Famille de densités GEP et p-crédence	71
2.6 Exemples	75
2.6.1 Student	79
2.6.2 GEP de type V	92
CHAPITRE III	

MÉTHODE D'ESTIMATION BAYÉSIENNE AVEC PARAMÈTRE D'ÉCHELLE ALÉATOIRE	103
3.1 Contexte théorique	103
3.2 Contexte bayésien	104
3.3 Le problème de position	104
3.4 Exemples	106
3.4.1 Student	109
3.4.2 GEP de type V	112
CONCLUSION	115
APPENDICE A PROGRAMMES INFORMATIQUES	118
RÉFÉRENCES	144

LISTE DES TABLEAUX

1.1	Coordonnées x et y pour les 10 points de l'ensemble de données	7
1.2	Répartition des observations selon la taille d'échantillon n lorsque toutes les coordonnées x_i des points P_i sont distinctes	50
1.3	Répartition des points selon le tableau 1.2 lorsque l'ordonnée du point P_{10} tend vers plus ou moins l'infini	53
1.4	Comportement de l'estimé de β pour les trois points sommaires différents, lorsque l'ordonnée du point P_{10} tend vers plus ou moins l'infini	54
1.5	Répartition des données originales selon la position de l'abscisse du point P_{10} par rapport aux autres coordonnées x_i , pour $i = 1, \dots, 9$	58
1.6	Valeurs sommaires des groupes G et D selon la valeur de la coordonnée x_{10} . Les valeurs sommaires de G et D sont définies par $\text{med}_{(x_i; y_i) \in G}(e_i^{(k)})$ et $\text{med}_{(x_i; y_i) \in D}(e_i^{(k)})$, respectivement	60
2.1	Coordonnées x et y pour les 10 points de l'ensemble de données	75
3.1	Coordonnées x et y pour les 10 points de l'ensemble de données	107
3.2	Estimés b et $\hat{\sigma}$ pour certaines valeurs de la coordonnée y_{10} ($z_{10} = y_{10}/x_{10}$), lorsque la variable $Z_i X_i, \beta, \sigma$ est distribuée selon une loi $t_4\left(\beta; \sigma\left(\frac{0.9106}{x_i}\right)\right)$	110
3.3	Estimés b et $\hat{\sigma}$ pour certaines valeurs de la coordonnée y_{10} ($z_{10} = y_{10}/x_{10}$), lorsque la variable $Z_i X_i, \beta, \sigma$ est distribuée selon une loi H_{GEP}	113

LISTE DES FIGURES

1.1	Les 10 points originaux avec la droite de régression obtenue par la méthode des moindres carrés. La pente de la droite est $b = 2,6269$	8
1.2	Estimé des moindres carrés de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	9
1.3	Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} .	10
1.4	Estimé des moindres carrés de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,1; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	11
1.5	Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} .	12
1.6	Comportement de l'estimé $\hat{\sigma}$ associé aux moindres carrés (a) pour différentes valeurs de la coordonnées y_{10} , (b) pour différentes valeurs de la coordonnée x_{10}	14
1.7	(a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés pour l'estimateur des moindres carrés. . . .	18
1.8	Fonction de poids reliée à la méthode itérative des moindres carrés repondérés pour l'estimateur des moindres carrés.	19
1.9	(a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés pour l'estimateur des moindres valeurs absolues.	21

1.10	Fonction de poids reliée à la méthode itérative des moindres carrés repondérés pour l'estimateur des moindres valeurs absolues.	22
1.11	Estimé des moindres valeurs absolues de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	23
1.12	Estimé des moindres valeurs absolues de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ lorsque le point $P_{10}^{(V)} = (7,1; 19,5)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	24
1.13	Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10}	25
1.14	Estimé des moindres valeurs absolues de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,7; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	26
1.15	Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10}	27
1.16	Comportement de l'estimé $\hat{\sigma}$ associé aux moindres valeurs absolues (a) pour différentes valeurs de la coordonnée y_{10} , (b) pour différentes valeurs de la coordonnée x_{10}	28
1.17	(a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés pour l'estimateur de Huber avec $m = 1,5$	29
1.18	Fonction de poids reliée à la méthode itérative des moindres carrés repondérés pour l'estimateur de Huber avec $m = 1,5$	30
1.19	Estimé de Huber de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	31

1.20	Comportement de l'estimé b de Huber pour (a) différentes valeurs de la coordonnée y_{10} (b) différentes valeurs de la coordonnée x_{10}	32
1.21	Estimé de Huber de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,7; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	33
1.22	L'estimé de b en fonction du nombre d'itérations effectuées avec la méthode des moindres carrés repondérés, lorsque $x_{10} = -29,3$	34
1.23	Comportement de l'estimé $\hat{\sigma}$ associé à Huber (a) pour différentes valeurs de la coordonnées y_{10} , (b) pour différentes valeurs de la coordonnée x_{10}	35
1.24	(a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés pour l'estimateur d'Andrews avec constante $A = 2,1$	36
1.25	Fonction de poids reliée à la méthode itérative des moindres carrés repondérés pour l'estimateur d'Andrews avec constante $A = 2,1$	37
1.26	Estimé d'Andrews de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	39
1.27	Comportement, pour différentes valeurs de la coordonnées y_{10} , (a) de l'estimé b et (b) de l'estimé $\hat{\sigma}$, tous deux associés à Andrews.	40
1.28	Estimé d'Andrews de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,7; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.	41
1.29	Comportement, pour différentes valeurs de la coordonnées x_{10} , (a) de l'estimé β et (b) de l'estimé $\hat{\sigma}$, tous deux associés à Andrews.	42

- 1.30 (a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés, pour l'estimateur bipoids de Tukey avec constante $B = 8$ 43
- 1.31 Fonction de poids reliée à la méthode itérative des moindres carrés repondérés, pour l'estimateur bipoids de Tukey avec constante $B = 8$. . . 44
- 1.32 Estimé bipoids de Tukey de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. 45
- 1.33 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} . 46
- 1.34 Estimé bipoids de Tukey de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,7; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. 47
- 1.35 Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} . 48
- 1.36 Comportement de l'estimé $\hat{\sigma}$ associé à Tukey (a) pour différentes valeurs de la coordonnée y_{10} , (b) pour différentes valeurs de la coordonnée x_{10} . 49
- 1.37 Estimé de la droite résistante de Tukey de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. 55
- 1.38 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} . 56
- 1.39 Estimé de la droite résistante de Tukey de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,1; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée x a été modifiée. 57
- 1.40 Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} . 58

- 1.41 Droites estimées par la méthode de la droite résistante pour le modèle $y_i = \beta x_i + \varepsilon_i$ (en noir) et pour le modèle $y_i = \alpha + \beta x_i + \varepsilon_i$ (en rouge). 62
- 2.1 Densités $N(0; 1)$ et F où F est une loi de Student avec 4 degrés de liberté et paramètre d'échelle $\tau_0 = 0,9106$ 81
- 2.2 Estimé bayésien de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. Les variables aléatoires Z_i suivent une loi $t_4(\beta; \frac{2,2765}{x_i})$ 82
- 2.3 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} . 83
- 2.4 Comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque (a) $y_{10} = 18,8$, (b) $y_{10} = 28,8$ et (c) $y_{10} = 68,8$. 84
- 2.5 Estimé bayésien de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,1; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. Les variables aléatoires Z_i suivent une loi $t_4(\beta; \frac{2,2765}{x_i})$ 86
- 2.6 Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} . 87
- 2.7 Comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque (a) $x_{10} = 17,1$, (b) $x_{10} = 57,1$, (c) $x_{10} = 407,1$ et (d) $x_{10} = 10007,1$ 89
- 2.8 Exemple de droites si le modèle rejette les points P_1, \dots, P_9 91
- 2.9 Densités $N(0; 1)$ et F , où F est une loi GEP de type V avec paramètres $\gamma = \delta = 0$, $\alpha = 1$, $\lambda = 4$, $z_0 = 2$ et paramètre d'échelle $\tau_0 = 1,7109$ 94

- 2.10 Estimé bayésien de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1 -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. Les variables aléatoires Z_i suivent une loi F_{GEP} 95
- 2.11 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} . 96
- 2.12 Comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque (a) $y_{10} = 18,8$, (b) $y_{10} = 28,8$ et (c) $y_{10} = 68,8$. 97
- 2.13 Estimé bayésien de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,1; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée x a été modifiée. Les variables aléatoires Z_i suivent une loi F_{GEP} 99
- 2.14 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} . 100
- 2.15 Comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque (a) $x_{10} = 57,1$, (b) $x_{10} = 1007,1$, (c) $x_{10} = 2007,1$ et (d) $x_{10} = 10007,1$ 102
- 3.1 Comportement de (a) l'estimé b et de (b) l'estimé $\hat{\sigma}$ pour différentes valeurs de la coordonnée y_{10} , lorsque la variable $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi $t_4 \left(\beta; \sigma \left(\frac{0,9106}{x_i} \right) \right)$ 111
- 3.2 Comportement de (a) l'estimé b et de (b) l'estimé $\hat{\sigma}$ pour différentes valeurs de la coordonnée y_{10} lorsque la variable $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi H_{GEP} 114

RÉSUMÉ

Dans ce mémoire nous nous intéressons à des méthodes d'estimations robustes de la pente de la droite de régression linéaire simple ainsi que du paramètre d'échelle de la densité des erreurs en présence de valeurs aberrantes dans l'échantillon de données. Une revue des méthodes d'estimations des paramètres de la droite de régression est présentée. Nous y analysons numériquement les différentes méthodes afin de décrire le comportement des estimateurs en présence d'une valeur aberrante dans l'échantillon. Une méthode d'estimation bayésienne est présentée afin d'estimer la pente de la droite de régression lorsque le paramètre d'échelle est connu. Nous exprimons le problème d'estimation de la pente de la droite de régression en un problème d'estimation d'un paramètre de position, ce qui nous permet d'utiliser les résultats de robustesse bayésienne pour un paramètre de position. Le comportement de cet estimateur est ensuite étudié numériquement lorsqu'il y a une valeur aberrante dans l'échantillon de données. Enfin, nous explorons une méthode bayésienne d'estimation simultanée du paramètre d'échelle et de la pente de la droite de régression. Nous exprimons le problème comme une estimation des paramètres de position et échelle même si les résultats de robustesse bayésienne pour ce cas ne sont pas encore publiés. Nous étudions tout de même le comportement des estimateurs de façon numérique.

MOTS CLÉS : Régression linéaire, inférence bayésienne, robustesse, valeurs aberrantes, densités à ailes relevées, densités GEP (*Generalized exponential power*), p-credence.

INTRODUCTION

La régression linéaire simple est un outil couramment utilisé dans plusieurs domaines avec le but de modéliser la relation existant entre deux variables d'intérêts. La technique consiste à faire passer une droite de façon à ce qu'elle se rapproche le plus possible de la tendance linéaire décrite par les données en estimant les paramètres de la droite. Toutefois, lorsque certaines observations dévient de façon marquée de la trajectoire linéaire, plusieurs méthodes d'estimations en sont affectées, produisant des droites plus ou moins représentatives des observations. Ces données, appelées valeurs aberrantes, proviennent souvent d'une saisie erronée des données ou d'une mauvaise technique d'échantillonnage, par exemple. Plusieurs méthodes d'estimations ont été développées afin de limiter ou de supprimer l'influence de telles valeurs. Lorsqu'elles y parviennent, on qualifie alors ces méthodes de robustes.

Dans ce mémoire, nous nous intéressons à la robustesse de méthodes d'estimations des paramètres de la régression linéaire, incluant le paramètre de la pente de la droite et le paramètre d'échelle de la densité des erreurs. Plus particulièrement, nous nous intéresserons à un modèle dont la droite passe par l'origine. Au chapitre 1, la robustesse d'estimateurs de la pente de la droite de régression et du paramètre d'échelle est discutée dans un contexte fréquentiste. À la section 1.2, nous débutons par l'analyse du cas classique de l'estimation par la méthode des moindres carrés, ce qui est équivalent à la méthode d'estimation du maximum de vraisemblance en supposant la densité des erreurs normale. La section 1.3 est consacrée à la classe d'estimateurs-M, une classe qui englobe entre autres la méthode des moindres carrés. L'algorithme du processus itératif permettant l'estimation des paramètres y est discuté. Nous analysons la robustesse de certains des estimateurs-M les plus répandus dans la littérature, tel que l'estimateur de Huber, Andrews, des moindres valeurs absolues et bipoids de Tukey. Nous terminons

le chapitre 1 par la présentation, à la section 1.4, de l'estimation par la méthode de la droite résistante de Tukey, suivi d'une analyse du comportement de cet estimateur en présence d'une valeur aberrante. Dans cette section, l'estimation est limitée à la pente de la droite de régression étant donné que cette méthode n'est pas conçue pour l'estimation d'un paramètre d'échelle.

Au chapitre 2, nous proposons une méthode d'estimation bayésienne robuste pour la pente de la droite de régression lorsque le paramètre d'échelle est considéré connu. Les premières sections de ce chapitre introduisent les différents contextes nécessaires à l'application de la théorie sur la robustesse présentée à la section 2.5. Ainsi, à la section 2.2, nous décrivons le contexte d'inférence bayésienne. Ensuite, les motivations nous amenant à exprimer le problème d'estimation de la pente de la droite de régression en un problème d'estimation d'un paramètre de position sont discutées à la section 2.3. Nous présentons dans la section 2.4 un survol des différents ouvrages sur l'estimation bayésienne robuste. La section 2.5 discute de la théorie sur la robustesse d'un paramètre de position dans un contexte bayésien. Entre autres, nous énumérons des conditions d'épaisseurs et de régularité que doivent satisfaire les ailes d'une densité à des fins de robustesse. Ces conditions sont ensuite simplifiées grâce à la notion de p -crédence à droite et à gauche, permettant de caractériser les ailes d'une densité. À la section 2.6, la robustesse de l'estimateur bayésien de la pente de la droite de régression est analysée en présence d'une valeur aberrante dans l'échantillon. Deux densités ayant des ailes plus relevées que celles de la loi normale sont considérées : la loi de Student et la loi GEP de type V.

Au chapitre 3, nous étudions la méthode d'estimation vue au chapitre précédent, mais en considérant le paramètre d'échelle inconnu et aléatoire. Nous exprimons la régression comme un problème de position et d'échelle. La contribution originale la plus importante de ce mémoire est d'avoir exprimé la régression, d'abord en un problème de position (avec échelle connue) puis en un problème de position et d'échelle. Nous pouvons ainsi analyser la robustesse de la régression à travers le prisme de la robustesse bayésienne et arriver ainsi à une régression bayésienne robuste remarquable. La théorie concernant

l'estimation simultanée de la pente de la droite de régression et du paramètre d'échelle étant en cours de développement, l'approche adoptée consiste à explorer numériquement le comportement des estimateurs en présence d'une valeur aberrante. Dans la section 3.2, nous présentons un contexte bayésien plus général que celui du chapitre 2. Le contenu théorique de la section 3.3 est similaire à celui de la section 2.3, mais en y ajoutant un paramètre d'échelle inconnu et aléatoire. Enfin, la section 3.4 explore de façon numérique la robustesse des estimateurs du paramètre de la régression et du paramètre d'échelle à l'aide des lois Student et GEP de type V.

CHAPITRE I

MÉTHODES DE RÉGRESSION ROBUSTE

1.1 Contexte théorique

La régression linéaire simple consiste à ajuster une droite le plus fidèlement possible à un ensemble de données $(x_i, y_i), i = 1, \dots, n$. Le modèle est défini par l'équation suivante :

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad (1.1)$$

où ε_i est un terme d'erreur. On suppose généralement que les erreurs sont des variables aléatoires d'espérance nulle. Afin d'utiliser le modèle, on estime les paramètres de la droite, c'est-à-dire l'ordonnée à l'origine α ainsi que la pente β , par a et b respectivement. La méthode d'estimation des moindres carrés (MC) est la plus couramment utilisée et ce, malgré qu'elle ait été développée au début des années 1800 (voir Rousseeuw et Leroy, 2003). Cependant, au fil du temps, son utilisation fréquente a mis à jour une de ses faiblesses. La méthode des moindres carrés n'est pas à l'abri d'une influence négative, voire catastrophique, causée par une ou plusieurs valeurs aberrantes sur les estimateurs des paramètres du modèle de régression. Cela se traduit généralement par un trop grand poids donné aux observations s'écartant significativement de la tendance générale de l'ensemble de points. On dit alors que la méthode n'est pas robuste aux valeurs aberrantes. Dans cette section, nous décrirons quelques méthodes d'estimations robustes. À l'aide d'un jeu de données, nous étudierons le comportement de ces méthodes lorsqu'une valeur est aberrante tant en abscisse qu'en ordonnée. Bien que l'équation (1.1)

contienne deux paramètres (α et β), nous mettrons l'emphase, dans ce mémoire, sur le modèle dont l'ordonnée à l'origine α est nulle, soit

$$y_i = \beta x_i + \varepsilon_i. \quad (1.2)$$

En effet, la méthode d'estimation robuste que nous proposons a d'abord été développée pour estimer le paramètre du modèle (1.2). Il existe plusieurs méthodes d'estimations robustes. Les méthodes qui seront décrites dans ce mémoire ne constituent pas une liste exhaustive des techniques existantes. Les méthodes choisies l'ont été de façon subjective et nous nous excusons auprès des auteurs des méthodes dont nous ne traiterons pas. Notons tout de même, parmi ces méthodes, celle de la droite de la moindre médiane («least median of squares», Rousseeuw, 1984) et de la droite la plus profonde («deepest line», Rousseeuw et Hubert, 1999). Dans le cas de la technique de la droite de la moindre médiane, notée LMS, elle consiste à minimiser, par rapport à α et β , l'expression suivante :

$$\text{med}_i \varepsilon_i^2 = \text{med}_i (y_i - \alpha - \beta x_i)^2,$$

pour $i = 1, \dots, n$ et où $\text{med}_i(h_i) = \text{médiane}(h_1, \dots, h_n)$.

Dans le cas de la technique de la droite la plus profonde, Rousseeuw et Huber définissent d'abord le vecteur de paramètres $\theta = (\alpha, \beta)$ comme un «nonfit» si et seulement si il existe un nombre v différent de x_i , tel que $\varepsilon_i < 0 \forall x_i < v$ et $\varepsilon_i > 0 \forall x_i > v$ ou $\varepsilon_i > 0 \forall x_i < v$ et $\varepsilon_i < 0 \forall x_i > v$, où $\varepsilon_i = y_i - \alpha - \beta x_i$, $i = 1, \dots, n$. Les estimateurs de α et β , par la technique de la droite la plus profonde, sont les valeurs de a et b qui maximisent en α et β

$$\max_{\alpha, \beta} rdepth(\varepsilon_i),$$

où $rdepth(\varepsilon_i)$ est le nombre de points minimal que l'on doit soustraire à l'ensemble de données afin que $\theta = (\alpha, \beta)$ soit un «nonfit».

1.2 Estimateur des moindres carrés

La technique des moindres carrés est la méthode d'estimation des paramètres la plus répandue, entre autres, à cause de la simplicité des calculs mathématiques, et la plus couramment implémentée dans les logiciels statistiques. Pour le modèle (1.1), l'estimation des moindres carrés consiste à minimiser, par rapport à α et β , l'expression suivante :

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2. \quad (1.3)$$

Les estimateurs de β et α , notés b et a , sont donnés par

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (1.4)$$

$$a = \bar{y} - b\bar{x},$$

où $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ et $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Les valeurs a et b obtenues sont, respectivement, des estimés de l'ordonnée à l'origine et de la pente de l'équation (1.1). Pour le cas particulier où α est nul, l'estimateur des moindres carrés du paramètre β de l'équation (1.2) est obtenu en minimisant l'équation

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta x_i)^2 \quad (1.5)$$

par rapport à β . L'estimateur du paramètre β est alors donné par l'équation

$$b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}. \quad (1.6)$$

Afin d'étudier le comportement de l'estimateur b , la méthode des moindres carrés est appliquée aux 10 points du tableau 1.1. Ce jeu de données a été simulé à partir du

modèle $y_i = 2,5x_i + \varepsilon_i$, avec $\varepsilon_i \sim N(0; 2,5^2)$. La robustesse de l'estimateur b pour le modèle (1.2) sera ensuite analysée en faisant varier un des points, soit en augmentant ou en diminuant d'abord son ordonnée, puis son abscisse. Cette donnée jouera le rôle de valeur aberrante.

i	x_i	y_i
1	1,1	2,4
2	1,2	0,5
3	1,5	7,6
4	3,0	10,3
5	3,7	6,4
6	3,9	9,2
7	3,9	10,7
8	5,0	15,6
9	5,7	13,6
10	7,1	18,8

Tableau 1.1 Coordonnées x et y pour les 10 points de l'ensemble de données

Notons chaque coordonnée par $P_i = (x_i; y_i)$, $i = 1, \dots, 10$. Sans perte de généralité, nous avons choisi de faire varier les coordonnées du point $P_{10} = (7,1; 18,8)$. Les dix points P_i originaux sont représentés dans la figure 1.1 avec la droite de régression obtenue par la méthode des moindres carrés. On y perçoit une tendance linéaire positive entre les variables x et y . L'estimation de β , par la méthode des moindres carrés, donne un estimé $b = 2,6269$. La droite estimée décrit la tendance linéaire de l'ensemble des points. Pour constater la faible influence du point P_{10} sur la pente de la droite de l'équation (1.2), l'estimation par la méthode des moindres carrés est refaite en enlevant le point P_{10} , ce qui donne $b_{MC} = 2,6178$, soit une légère diminution par rapport à $b = 2,6269$. Si le point P_{10} n'a que peu d'influence sur l'estimation de la pente β , il en va autrement de la variance de b . Ainsi, dans ce mémoire, lorsqu'il est question de l'influence du point P_{10} sur la pente, nous ne considérons que l'influence qu'à le point sur la valeur de b .

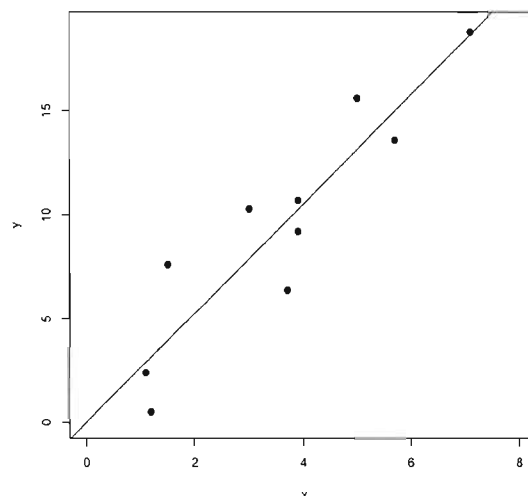


Figure 1.1 Les 10 points originaux avec la droite de régression obtenue par la méthode des moindres carrés. La pente de la droite est $b = 2,6269$.

Considérons d'abord le cas où le point P_{10} est clairement aberrant en ordonnée, en considérant les nouveaux points $P_{10}^{(I)} = (7,1; 68,8)$ et $P_{10}^{(II)} = (7,1; -31,2)$, soit le point P_{10} auquel on a ajouté et soustrait 50 à y_{10} . Les points $P_{10}^{(I)}$ et $P_{10}^{(II)}$ s'écartent considérablement de la tendance linéaire du reste des points, comme on peut le voir sur les graphiques (a) et (b) de la figure 1.2. Les estimés de β pour chaque graphique sont donnés respectivement par $b_I = 4,7662$ et $b_{II} = 0,4872$. Les points $P_{10}^{(I)}$ et $P_{10}^{(II)}$ attirent la droite de façon importante. Ils ont une grande influence sur la droite par rapport aux autres points. Cela force la droite à faire un mauvais compromis entre la valeur aberrante et les autres points, dans la mesure où la droite passe par une région dans laquelle ne se trouve presque aucun point. La méthode n'est clairement pas robuste.

Notons que nous n'aborderons pas les techniques permettant de déterminer à quel moment une donnée est considérée aberrante. Pour qu'une méthode d'estimation soit dite robuste, on s'attend à ce que l'influence du point P_{10} sur la pente β soit limitée, ou mieux,

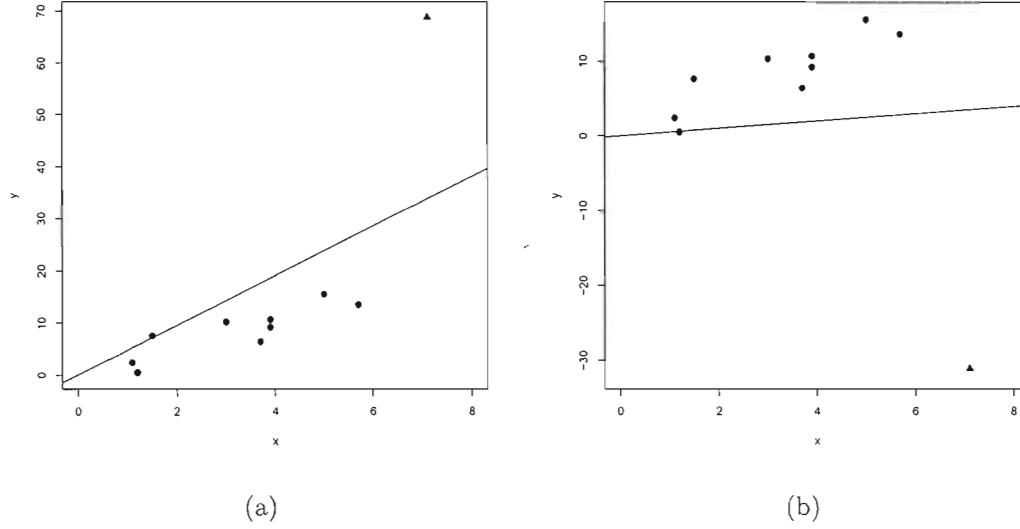


Figure 1.2 Estimé des moindres carrés de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7, 1; 68, 8)$, (b) lorsque le point $P_{10}^{(II)} = (7, 1; -31, 2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

qu'elle s'estompe à mesure que le point s'éloigne du reste de l'ensemble de points. Ce n'est pas le cas pour la méthode d'estimation des moindres carrés, comme on peut le voir dans la figure 1.3. En effet, plus la valeur de y_{10} augmente, plus la valeur de b augmente et vice-versa. Le comportement de b peut s'expliquer par le fait que, de l'équation (1.6), on peut exprimer b comme une fonction linéaire en y_{10} , c'est-à-dire $b = c_1 + c_2 y_{10}$ où

$$c_1 = \frac{\sum_{i=1}^9 x_i y_i}{\sum_{i=1}^{10} x_i^2} \quad \text{et} \quad c_2 = \frac{x_{10}}{\sum_{i=1}^{10} x_i^2}.$$

En effet, de l'équation (1.6), pour $n = 10$, on a

$$\begin{aligned} b &= \frac{\sum_{i=1}^{10} x_i y_i}{\sum_{i=1}^{10} x_i^2} \\ &= \frac{\sum_{i=1}^9 x_i y_i}{\sum_{i=1}^{10} x_i^2} + \frac{x_{10}}{\sum_{i=1}^{10} x_i^2} y_{10} \\ &= 1,822374 + 0,0427943 y_{10}. \end{aligned}$$

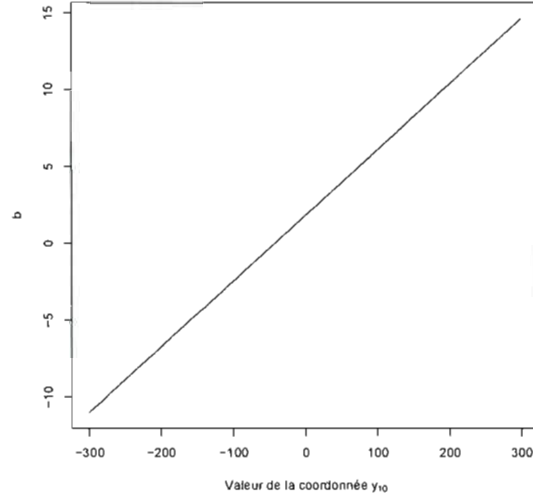


Figure 1.3 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} .

La valeur de $c_2 = 0,0427943$ correspond exactement à la pente de la droite de la figure 1.3.

Considérons maintenant le cas où le point P_{10} est clairement aberrant en abscisse, c'est-à-dire le point P_{10} dont on a ajouté et enlevé 50 à la coordonnée x_{10} . Les graphiques (a) et (b) de la figure 1.4 nous indiquent clairement que les points $P_{10}^{(III)} = (57,1; 18,8)$ et $P_{10}^{(IV)} = (-42,9; 18,8)$ influencent la trajectoire de la droite en l'attirant à mesure qu'ils s'éloignent des autres points. Les estimés de β sont donnés respectivement par $b_{III} = 0,4075$ et $b_{IV} = -0,2578$. Encore une fois, la droite se compromet en passant par une région dans laquelle ne se trouve presque aucun point. D'ailleurs, sur le graphique (b) de la figure 1.4, on constate que la droite passe sous tous les points (elle y est forcée puisqu'elle doit passer par l'origine), ce qui s'avère un comportement peu souhaitable si on désire modéliser l'ensemble des points. Le fait que les estimés b_{III} et b_{IV} soient près de 0 est illustré par la figure 1.5. Celle-ci nous indique que lorsque l'abscisse du point P_{10} tend vers plus ou moins l'infini, la pente de la droite de régression converge vers

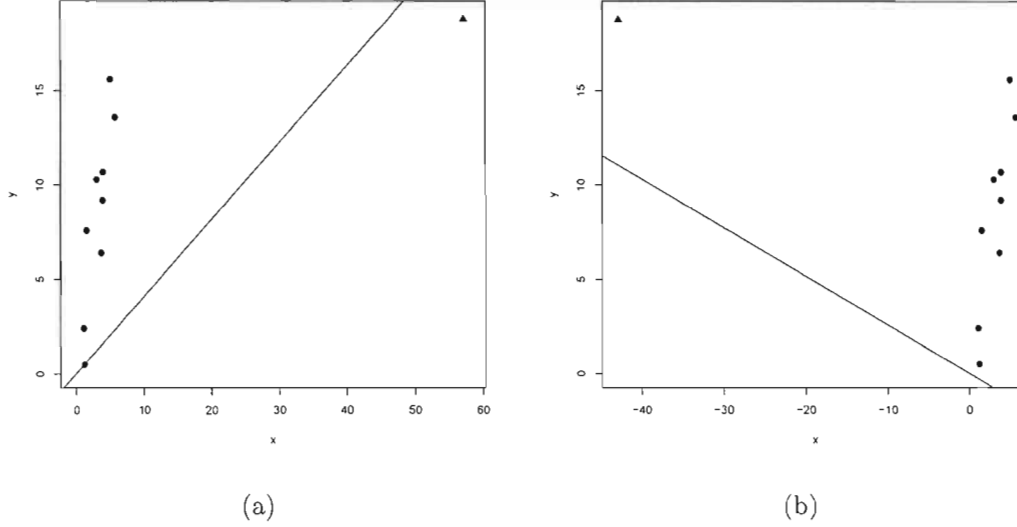


Figure 1.4 Estimé des moindres carrés de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,1; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole ▲ représente le point P_{10} dont la coordonnée y a été modifiée.

0. En considérant l'équation (1.6), le comportement de l'estimé b peut s'expliquer de la façon suivante :

$$\begin{aligned}
 \lim_{|x_{10}| \rightarrow \infty} b &= \lim_{|x_{10}| \rightarrow \infty} \frac{\sum_{i=1}^{10} x_i y_i}{\sum_{i=1}^{10} x_i^2} \\
 &= \lim_{|x_{10}| \rightarrow \infty} \left(\frac{\sum_{i=1}^9 x_i y_i}{\sum_{i=1}^{10} x_i^2} + \frac{x_{10} y_{10}}{\sum_{i=1}^{10} x_i^2} \right) \\
 &= \lim_{|x_{10}| \rightarrow \infty} \left(\frac{\sum_{i=1}^9 x_i y_i}{\sum_{i=1}^{10} x_i^2} + \frac{y_{10}}{\sum_{i=1}^{10} x_i^2 / x_{10}} \right) \\
 &= \lim_{|x_{10}| \rightarrow \infty} \left(\frac{\sum_{i=1}^9 x_i y_i}{\sum_{i=1}^9 x_i^2 + x_{10}^2} + \frac{y_{10}}{\sum_{i=1}^9 x_i^2 / x_{10} + x_{10}} \right) \\
 &= 0.
 \end{aligned}$$

On remarque que lorsque x_{10} est très grand, $b \approx \frac{y_{10}}{x_{10}}$. Dans ce cas, c'est le point $P_{10} = (x_{10}; y_{10})$ qui détient pratiquement tout le poids sur la droite ancrée à l'origine. On constate que la méthode d'estimation des moindres carrés n'est pas robuste en présence

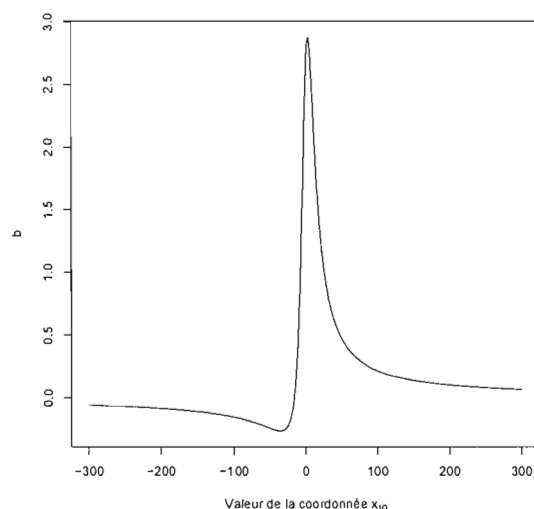


Figure 1.5 Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} .

d'une valeur aberrante en x .

Des hypothèses additionnelles sur les erreurs ε_i sont généralement posées dans un modèle de régression linéaire. Les erreurs ε_i sont réputées être des variables aléatoires identiquement et indépendamment distribuées partageant le même paramètre d'échelle σ . On écrira $\frac{\varepsilon_i}{\sigma}$ i.i.d F , pour une distribution quelconque F . Si par exemple F est une $N(0, 1)$, cela revient à dire que ε_i i.i.d $N(0, \sigma^2)$. Notons que σ^2 représente dans ce cas particulier la variance et on a donc homoscédasticité. Généralement, le paramètre σ^2 est inconnu et doit être lui aussi estimé à partir de l'échantillon. Son estimé est nécessaire pour faire de l'inférence statistique sur le paramètre β . Nous utiliserons pour ce faire la méthode du maximum de vraisemblance, qui est une technique d'estimation qui permet d'estimer à la fois les paramètres β et σ^2 . Lorsque F est une $N(0, 1)$, on peut montrer que les méthodes des moindres carrés et du maximum de vraisemblance donnent le même estimé b . Pour le paramètre σ^2 , la méthode du maximum de vraisemblance donne

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - bx_i)^2.$$

Cependant, on utilisera généralement l'estimateur sans biais de σ^2 suivant :

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - bx_i)^2.$$

Comme σ^2 ne représente généralement pas la variance des erreurs ε_i , mais plutôt le carré du paramètre d'échelle, nous travaillerons avec le paramètre σ , estimé par

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - bx_i)^2}.$$

Dans ce mémoire, nous nous intéresserons aussi à la robustesse de $\hat{\sigma}$. Les graphiques (a) et (b) de la figure 1.6 illustrent le comportement de l'estimé $\hat{\sigma}$ lorsque le point P_{10} tend vers plus ou moins l'infini en ordonnée et en abscisse, respectivement. Cela nous permet d'avoir une idée de l'influence du point aberrant sur $\hat{\sigma}$. Le graphique (a) démontre clairement que $\hat{\sigma}$ n'est pas un estimateur robuste à une valeur aberrante en y . En effet, plus le point P_{10} s'éloigne vers plus ou moins l'infini, plus l'estimé $\hat{\sigma}$ augmente. Rappelons que le paramètre σ est une mesure de dispersion des erreurs, c'est-à-dire de la distance verticale entre les points et la droite. On voit alors que le graphique (a) de la figure 1.6 est cohérent avec le graphique (a) de la figure 1.2 et la figure 1.3. Pour le graphique (b), à mesure que le point P_{10} tend vers plus ou moins l'infini, l'influence qu'il exerce sur l'estimateur de σ se stabilise. Cela s'explique par l'estimé de la pente de la droite qui converge vers 0 (voir figure 1.5). L'influence de la valeur aberrante est donc limitée. Cela indique généralement une certaine robustesse. Cependant, l'estimateur de σ converge vers une valeur donnée par $\hat{\sigma} = 9,1557$, qui n'apparaît pas appropriée pour le jeu de données du tableau 1.1. En effet, la valeur de $\hat{\sigma}$ lorsque le point P_{10} est retiré est $\hat{\sigma}^* = 2,2802$, qui est nettement inférieur à 9,1557. On conclut que l'estimateur du

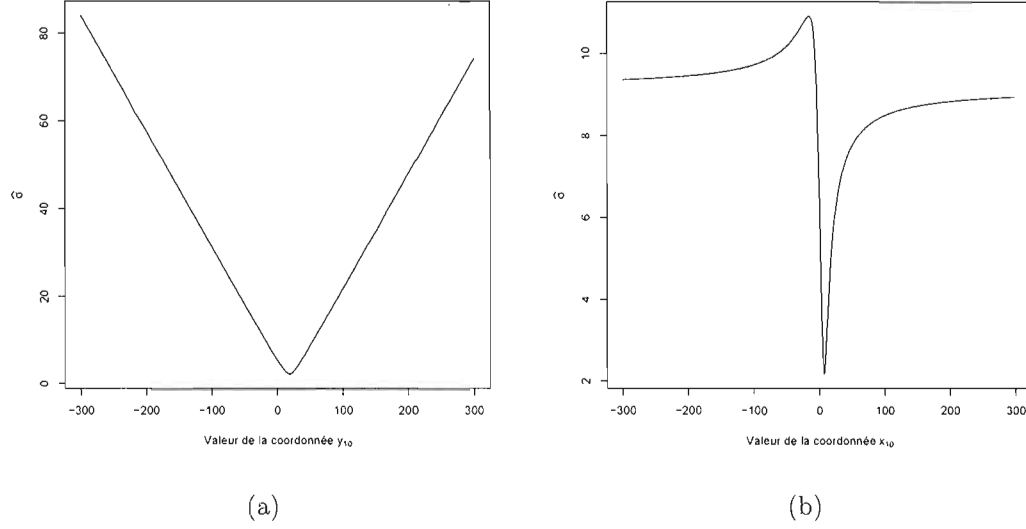


Figure 1.6 Comportement de l'estimé $\hat{\sigma}$ associé aux moindres carrés (a) pour différentes valeurs de la coordonnées y_{10} , (b) pour différentes valeurs de la coordonnée x_{10} .

paramètre d'échelle n'est pas robuste à une valeur aberrante en x , même si l'influence de la valeur aberrante est limitée.

1.3 Estimateurs-M

L'équation (1.5) peut être généralisée dans le but d'accroître la robustesse de l'estimateur b lorsqu'une ou plusieurs valeurs extrêmes se retrouvent dans l'échantillon. Huber (1964) a introduit la classe d'estimateurs-M en proposant l'idée suivante. Dans l'équation (1.5), Huber a remplacé ε_i^2 par une fonction des erreurs ε_i et de σ , notée $\rho\left(\frac{\varepsilon_i}{\sigma}\right)$, pour $i = 1, \dots, n$. La fonction $\rho(\cdot)$, appelée fonction objectif, est choisie de façon à être symétrique ($\rho(-u) = \rho(u)$) et à avoir un unique minimum en 0.

Pour le modèle défini à l'équation (1.2), on suppose que $\frac{\varepsilon_i}{\sigma}$, pour $i = 1, \dots, n$, sont des variables aléatoires indépendantes de même loi, notée F . Précisons qu'on ne pose pas d'hypothèse quant à la nature de la loi F . Le paramètre σ joue le rôle de paramètre

d'échelle.

L'estimateur-M de β est défini de la façon suivante :

Définition 1. Soit un modèle de régression linéaire simple défini par l'équation $y_i = \beta x_i + \varepsilon_i$, où $\frac{\varepsilon_i}{\sigma}$ i.i.d. selon une loi quelconque, pour $i = 1, \dots, n$. Alors b est un estimateur-M de β , si b minimise

$$\sum_{i=1}^n \rho\left(\frac{\varepsilon_i}{\sigma}\right) = \sum_{i=1}^n \rho\left(\frac{y_i - \beta x_i}{\sigma}\right). \quad (1.7)$$

Si on suppose que ρ est une fonction continue et dérivable en tout point, alors il s'avère souvent plus facile de chercher l'estimateur b qui est solution en β de

$$\sum_{i=1}^n \psi\left(\frac{y_i - \beta x_i}{\sigma}\right) x_i = 0, \quad (1.8)$$

où $\psi(u) = \frac{d}{d\theta} \rho(\theta)|_{\theta=u}$.

L'estimateur des moindres carrés est un cas particulier de l'estimateur-M en posant $\rho(u) = u^2$. Notons que dans ce cas particulier, le paramètre σ n'intervient pas dans le calcul de b . En effet, l'estimateur des moindres carrés de β est la valeur qui minimise

$$\begin{aligned} \sum_{i=1}^n \rho\left(\frac{\varepsilon_i}{\sigma}\right) &= \sum_{i=1}^n \left(\frac{\varepsilon_i}{\sigma}\right)^2 \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n \varepsilon_i^2 \\ &\propto \sum_{i=1}^n \varepsilon_i^2. \end{aligned}$$

On voit par la dernière ligne, que la minimisation en β ne dépend pas de σ .

1.3.1 Méthode itérative

Dans son ouvrage, Huber (1981) présente quelques méthodes permettant d'estimer β et σ , pour toute fonction objectif $\rho(\cdot)$. Il s'agit des méthodes itératives de Newton, de Huber et des moindres carrés repondérés (*reweighted least squares*). Ces méthodes se simplifient dans certains cas particuliers, par exemple lorsque l'estimateur de β ne requiert pas la connaissance du paramètre σ . Ces méthodes consistent à estimer β et σ à tour de rôle à partir d'une estimation préalable de β . Pour la simplicité de ses calculs mathématiques et son efficacité dans la compilation informatique, nous présentons uniquement la méthode itérative des moindres carrés repondérés. Pour les méthodes de Newton et Huber, voir Dutter (1977), Huber (1981), Hoaglin et al. (1983) et Hoaglin et al. (1985).

Méthode itérative des moindres carrés repondérés. L'intérêt de cette méthode est de pouvoir utiliser les techniques de calculs bien connues de la méthode des moindres carrés. Définissons d'abord la fonction de poids $w(u)$ par $w(u) = \frac{\psi(u)}{u}$. Le procédé itératif associe à chaque itération, et ce pour chaque point de l'ensemble, un poids calculé à partir de la fonction de poids $w(u)$. Ainsi, la méthode consiste à choisir d'abord un estimé initial pour β , que l'on note $b^{(0)}$. La nature de ce $b^{(0)}$ dépend de l'estimateur-M employé. Par la suite, on applique le procédé itératif de telle sorte qu'à l'itération $k + 1$, à partir de $b^{(k)}$ connu, on estime

$$e_i^{(k)} = y_i - b^{(k)} x_i,$$

$$\hat{\sigma}^{(k)} = \frac{1}{0,6745} \operatorname{med}_i \left[\left| e_i^{(k)} - \operatorname{med}_j \left[e_j^{(k)} \right] \right| \right], \quad (1.9)$$

$$w_i^{(k)} = w \left(\frac{e_i^{(k)}}{\hat{\sigma}^{(k)}} \right) = \frac{\psi \left[\frac{e_i^{(k)}}{\hat{\sigma}^{(k)}} \right]}{\left(\frac{e_i^{(k)}}{\hat{\sigma}^{(k)}} \right)} \quad (1.10)$$

$$b^{(k+1)} = \frac{\sum_{i=1}^n x_i y_i w_i^{(k)}}{\sum_{i=1}^n x_i^2 w_i^{(k)}},$$

où $\operatorname{med}_j(h_j) = \text{médiane}(h_1, \dots, h_n)$, pour une fonction quelconque h et où $w_i^{(k)}$ est une pondération qui doit être estimée à chaque itération. Notons que la valeur du paramètre d'échelle σ n'a de sens que si elle est associée à une distribution F spécifiée. Rappelons que $\frac{\epsilon_i}{\sigma}$ i.i.d F . Par exemple, pour un même jeu de données, un paramètre σ associé à une distribution F qui est $N(0, 100)$ sera dix fois plus petit qu'un σ associé à une $N(0, 1)$. Il peut sembler étonnant que l'estimation de σ donnée par l'équation (1.9) ne soit associée à aucune distribution F spécifiée. En fait, l'équation (1.9) est du même ordre que l'écart-type, en plus robuste. Cela signifie qu'implicitement, la distribution F associée est supposée avoir une dispersion semblable à la $N(0, 1)$, même si elle n'est pas spécifiée.

1.3.2 Exemples d'estimateurs-M

Nous allons décrire dans cette section quelques estimateurs-M souvent mentionnés dans les ouvrages sur la robustesse de l'estimation des paramètres de la régression. Il s'agit des estimateurs des moindres valeurs absolues, de Huber, d'Andrews et de Tukey. Pour chaque estimateur, nous utiliserons les données introduites à la section 1.2 afin d'analyser

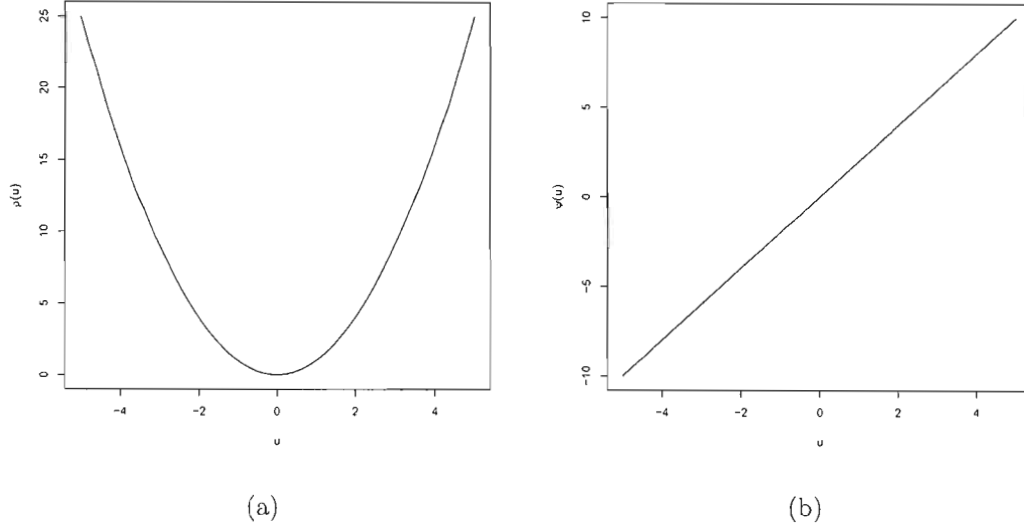


Figure 1.7 (a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés pour l'estimateur des moindres carrés.

son comportement en présence d'une valeur extrême. Il est important de choisir une valeur initiale de l'estimateur qui est adéquate pour s'assurer que le processus itératif converge et afin de diminuer le nombre d'itérations nécessaire.

Estimateur des moindres carrés

Comme nous l'avons vu au début de la section 1.3, l'estimateur de β obtenu par la technique des moindres carrés est un estimateur-M. Nous avons aussi vu que le calcul de b ne requiert pas la connaissance du paramètre d'échelle σ . La fonction objectif est donnée par $\rho(u) = u^2$ alors que $\psi(u) = 2u$ et $w(u) = 2$. On peut observer graphiquement ces fonctions sur les figures 1.7 (a) et (b) et 1.8, respectivement. Le processus itératif des moindres carrés repondérés converge en une itération car il accorde à chaque point de l'ensemble, un poids de $w_i^{(k)} = w\left(\frac{e_i^{(k)}}{\hat{\sigma}^{(k)}}\right) = 2$, ce qui résulte en $b^{(k+1)} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$, pour tout k . Le comportement de cet estimateur a déjà été étudié.

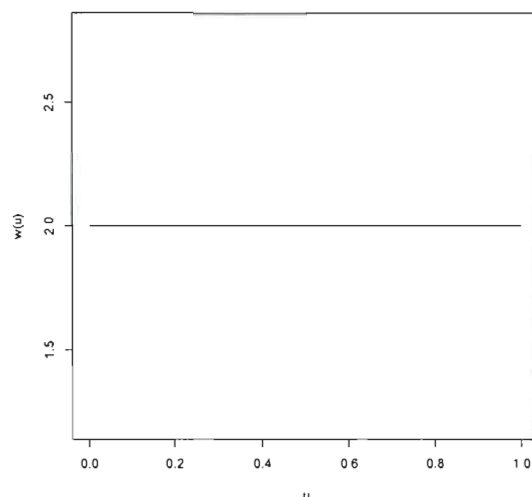


Figure 1.8 Fonction de poids reliée à la méthode itérative des moindres carrés pondérés pour l'estimateur des moindres carrés.

Estimateur des moindres valeurs absolues

La méthode d'estimation des moindres valeurs absolues (aussi appelée méthode L_1) a été développée vers le milieu des années 1700 par Roger Joseph Boscovich (voir Birkes et Dodge, 1993). Toutefois, la difficulté reliée aux calculs d'estimation en a fait une méthode peu populaire qui a rapidement perdu de son attrait lorsque la méthode des MC a vu le jour. Ce n'est que beaucoup plus tard que Francis Ysidro Edgeworth, reprenant les travaux de Boscovich, a constaté qu'il s'agissait d'une technique d'estimation pouvant atténuer le problème du trop grand poids attribué aux grandes valeurs résiduelles par la méthode des moindres carrés (voir Rousseeuw et Leroy 2003). La méthode L_1 consiste à trouver la valeur b qui minimise en β la fonction suivante :

$$\sum_{i=1}^n \left| \frac{\varepsilon_i}{\sigma} \right| = \sum_{i=1}^n \left| \frac{y_i - \beta x_i}{\sigma} \right| \propto \sum_{i=1}^n |y_i - \beta x_i|.$$

La fonction objectif est $\rho(u) = |u|$, alors que la fonction ψ est donnée par

$$\psi(u) = \text{sgn}(u) = \begin{cases} +1, & \text{si } u > 0; \\ 0, & \text{si } u = 0; \\ -1, & \text{si } u < 0. \end{cases}$$

Bien que la dérivée de la fonction objectif n'est pas définie en $u = 0$, on admet tout de même que $\psi(0) = 0$, soit le point milieu entre $\psi(0^-)$ et $\psi(0^+)$. On voit que $w(u) = \frac{1}{|u|}$ si $u \neq 0$ avec $\lim_{u \rightarrow 0} w(u) = \infty$. Tout comme dans le cas de la méthode des moindres carrés, l'estimation du paramètre β ne requiert pas la connaissance du paramètre σ . Le poids de l'équation (1.10) associé à l'estimateur L_1 , tel que proposé par Hoaglin et al. (1983), est donné par

$$w_i^{(k)} = w\left(\frac{e_i^{(k)}}{\hat{\sigma}^{(k)}}\right) = \frac{1}{(|e_i^{(k)}|/\hat{\sigma}^{(k)})},$$

si $\frac{|e_i^{(k)}|}{\hat{\sigma}^{(k)}} > 0$, avec $\lim_{|e_i^{(k)}| \rightarrow 0} w_i^{(k)} = \infty$. On a $\hat{\sigma}^{(k)}$ qui est constant par rapport à i et s'annulera dans le calcul de $b^{(k+1)}$. Les figures 1.9 (a) et (b) et 1.10 illustrent les fonctions ρ , ψ et w , respectivement. Afin d'étudier l'effet d'une valeur aberrante sur l'estimateur L_1 , nous allons à nouveau considérer les données du tableau 1.1, avec le point P_{10} auquel on a ajouté et retranché 50 à son ordonnée puis à son abscisse. Hoaglin et al. (1985) proposent d'utiliser comme valeur initiale $b^{(0)}$, l'estimé des MC obtenu de l'équation (1.6), soit $b^{(0)} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$.

Nous supposons maintenant que le procédé itératif a convergé pour un niveau de précision donné à l'itération k . Dans le but d'alléger la notation, e_i , $\hat{\sigma}$ et w_i sans l'indice k indiquent les valeurs lorsque le processus itératif a convergé. Les graphiques (a) et (b) de la figure 1.11 illustrent les cas où les point $P_{10}^{(I)} = (7,1; 68,8)$ et $P_{10}^{(II)} = (7,1; -31,2)$, respectivement, sont aberrants en ordonnée. Les estimés de β sont donnés respective-

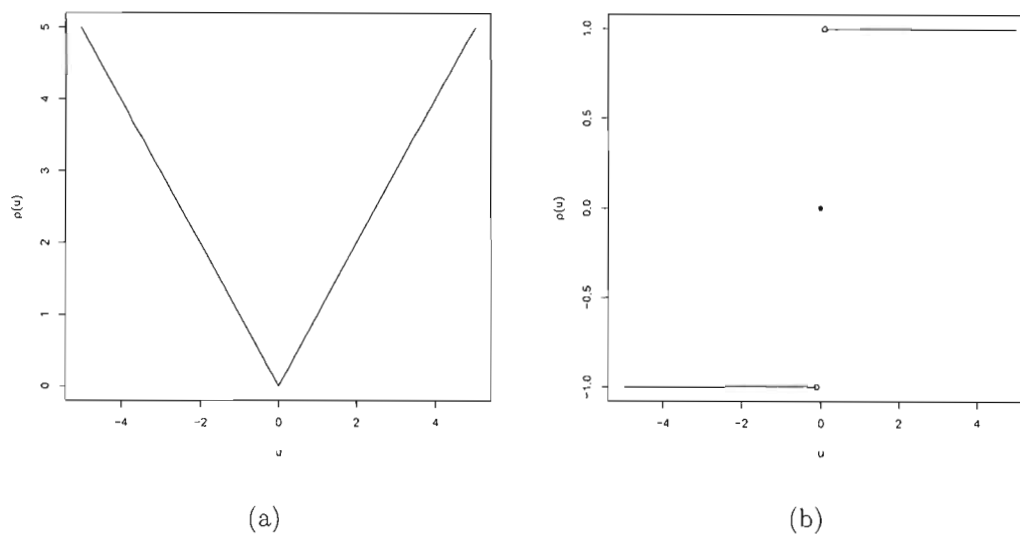


Figure 1.9 (a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés pour l'estimateur des moindres valeurs absolues.

ment par $b_I = 2,7436$ et $b_{II} = 2,3860$. Dans les deux cas, on remarque que la droite ne subit que très peu l'influence du point extrême. L'estimé de la pente s'éloigne peu de la valeur de $b_{L_1} = 2,3859$ qui correspond à l'estimé L_1 de β lorsque le point P_{10} est retiré des données. La méthode itérative des moindres carrés repondérés donne pratiquement tout le poids au point P_7 (ou P_9) lorsque l'ordonnée y_{10} tend vers plus (ou moins) l'infini, respectivement. En effet, on observe que $w_7 = \frac{\hat{\sigma}}{|e_7|}$ (ou $w_9 = \frac{\hat{\sigma}}{|e_9|}$) est très grand par rapport aux autres poids. Une particularité de la méthode L_1 est qu'elle produit un estimé b qui passe toujours par au moins un point de l'ensemble de données lorsque le modèle 1.2 est considéré (voir Birkes et Dodge, 1993). Puisque la droite doit passer par l'origine, on remarque que les estimés b_I et b_{II} sont déterminés uniquement par les points P_7 et P_9 respectivement. En effet, $b_I = \frac{y_7 - 0}{x_7 - 0} = \frac{10,7}{3,9} = 2,7436$ alors que

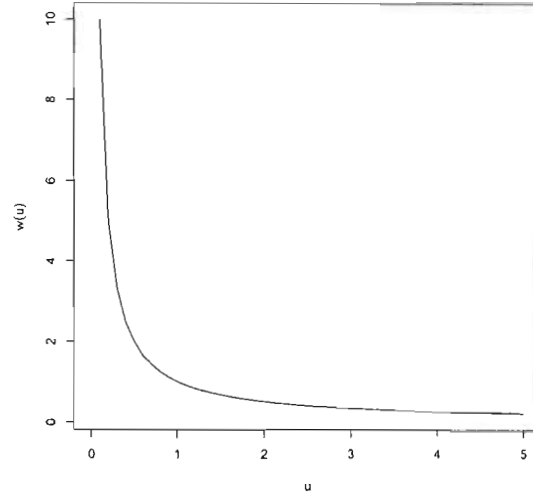


Figure 1.10 Fonction de poids reliée à la méthode itérative des moindres carrés pondérés pour l'estimateur des moindres valeurs absolues.

$b_{II} = \frac{y_9 - 0}{x_9 - 0} = \frac{13,6}{5,7} = 2,3860$. Ce qui est équivalent à

$$\begin{aligned} b_I &= \lim_{|w_7| \rightarrow \infty} \frac{\sum_{i=1}^{10} x_i y_i w_i}{\sum_{i=1}^{10} x_i^2 w_i} \\ &= \frac{x_7 y_7 w_7}{x_7^2 w_7} \\ &= \frac{y_7}{x_7} \end{aligned}$$

et

$$\begin{aligned} b_{II} &= \lim_{|w_9| \rightarrow \infty} \frac{\sum_{i=1}^{10} x_i y_i w_i}{\sum_{i=1}^{10} x_i^2 w_i} \\ &= \frac{y_9}{x_9}. \end{aligned}$$

Le point P_{10} a une influence importante sur l'estimation de β lorsque y_{10} se retrouve dans l'intervalle $[16,95; 19,45]$. Dans cet intervalle, la droite passe exactement par le point P_{10} . À mesure que le point P_{10} augmente entre 16,95 et 19,45, la pente de la droite progresse. On perçoit exactement cette situation sur la figure 1.13. La méthode

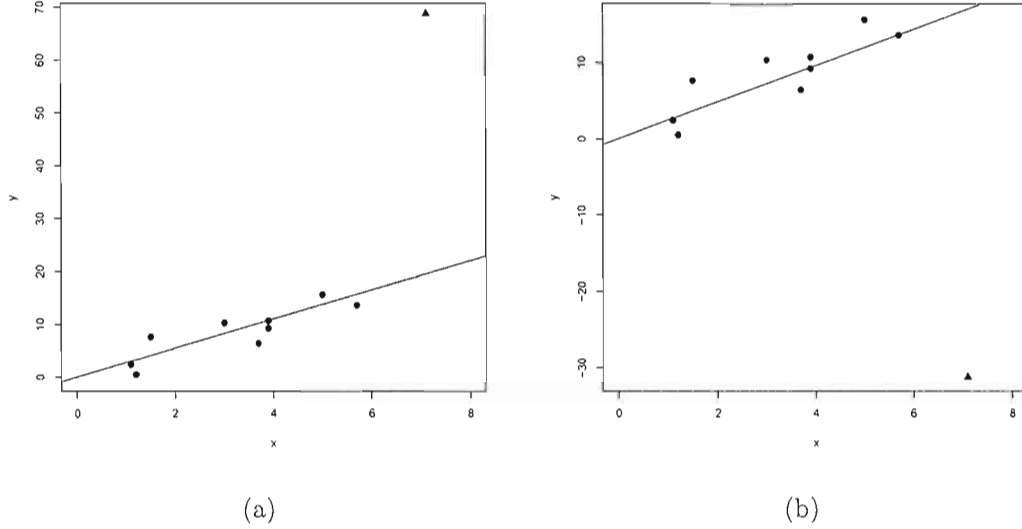


Figure 1.11 Estimé des moindres valeurs absolues de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

réagit de façon robuste à une valeur aberrante en y . Par contre, on s'attend à ce que le poids accordé à un point par une méthode d'estimation performante soit plus élevé si ce point est non aberrant que s'il est aberrant. À la figure 1.12, on considère le point $P_{10}^{(V)} = (7,1; 19,5)$, qui n'est pas une donnée aberrante par rapport aux neuf autres points P_i . Toutefois, le modèle n'y accorde pas plus de poids qu'il ne le faisait pour le point $P_{10}^{(I)} = (7,1; 68,8)$ du graphique (a) de la figure 1.11. En effet, l'estimé de β est le même dans les deux cas, $b_V = b_I = 2,7436$. Le modèle de régression L_1 est robuste aux valeurs aberrantes en ordonnée, mais manque de finesse lorsqu'il n'y a pas de points aberrants.

Le cas où le point P_{10} est aberrant en abscisse est représenté par les graphiques (a) et (b) de la figure 1.14. La pente de la droite pour chaque graphique est estimée à $b_{III} = 0,3293$ et $b_{IV} = -0,4382$ respectivement. Puisque le modèle force la droite à passer par l'origine, on observe que les estimés b_{III} et b_{IV} sont déterminés uniquement par les points $P_{10}^{(III)} = (57,1; 18,8)$ et $P_{10}^{(IV)} = (-42,9; 18,8)$ respectivement. Rappelons

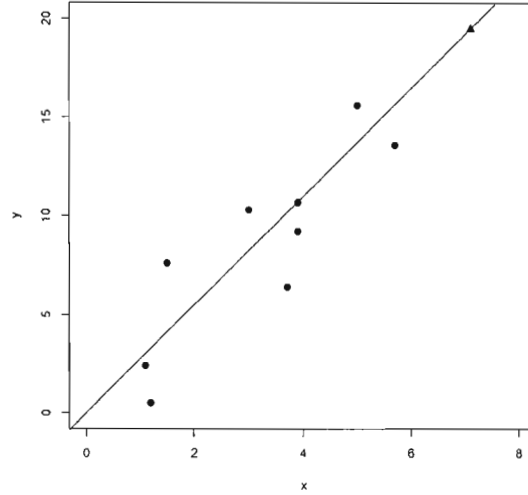


Figure 1.12 Estimé des moindres valeurs absolues de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ lorsque le point $P_{10}^{(V)} = (7, 1; 19, 5)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

que les erreurs représentent la distance verticale entre un point et la droite. On voit que si la droite passait autour des neuf autres points, l'erreur générée par la valeur aberrante serait gigantesque. Puisqu'on désire minimiser la somme des valeurs absolues des erreurs, on préférera rejeter ces neuf points et passer par le point $P_{10}^{(III)}$. Cela explique que la méthode L_1 ne produit pas un estimé b robuste à une valeur aberrante en x .

La figure 1.15 illustre le comportement de l'estimateur de β . Lorsque x_{10} tend vers plus ou moins l'infini, la valeur de l'estimé b tend vers zéro. Lorsque $|x_{10}|$ sera suffisamment grand, l'estimateur de β sera déterminé uniquement par le point $(x_{10}; 18, 8)$ et l'origine. La valeur de l'estimé sera alors

$$b = \lim_{|x_{10}| \rightarrow \infty} \frac{18,8 - 0}{x_{10} - 0} = 0.$$

Lorsque la valeur x_{10} se déplace dans la région des neuf autres points, la droite sautera d'un point à l'autre, ce qui explique les sauts de la figure (1.15).

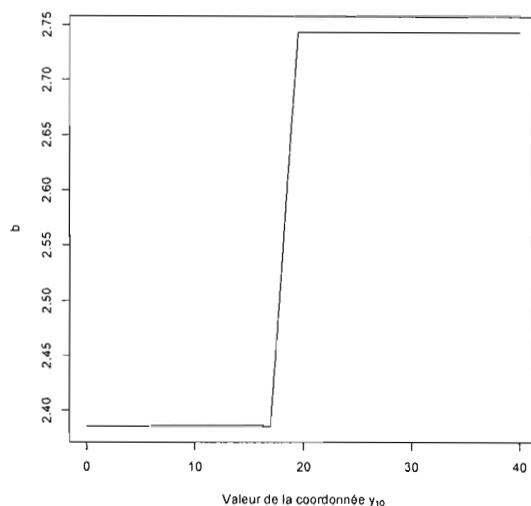


Figure 1.13 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} .

Nous nous intéressons maintenant au comportement de l'estimateur de σ de l'équation (1.9) face à un point aberrant en ordonnée et en abscisse. L'estimé de σ , lorsque le point P_{10} est supprimé de l'ensemble, est donné par $\hat{\sigma}_{L_1} = 3,5036$. Le graphique (a) de la figure 1.16 illustre le comportement de $\hat{\sigma}$. Lorsque le point P_{10} tend vers plus ou moins l'infini en ordonnée, l'estimé de σ devient constant. Cela s'explique par le fait que l'estimé de β devient lui aussi constant lorsque le point P_{10} tend vers plus ou moins l'infini (voir figure 1.13) et par la résistance qu'offre la médiane par rapport à des valeurs trop éloignées dans le calcul de $\hat{\sigma}$. Lorsque l'ordonnée y_{10} tend vers plus ou moins l'infini, l'estimé de σ est égal à 3,3872 ou 3,4737, respectivement. L'estimé sous-estime quelque peu $\hat{\sigma}_{L_1}$. L'estimateur $\hat{\sigma}$ est robuste à une valeur aberrante en y , mais tout comme pour l'estimateur de β , il manque de finesse.

Sur la figure 1.16 (b), lorsque le point P_{10} tend vers plus ou moins l'infini en abscisse, on observe que l'estimé de σ excède la valeur de $\hat{\sigma}_{L_1} = 3,5036$. On peut en déduire sur les figures 1.14 (a) et (b) que l'amplitude de la plupart des résidus est plus importante que

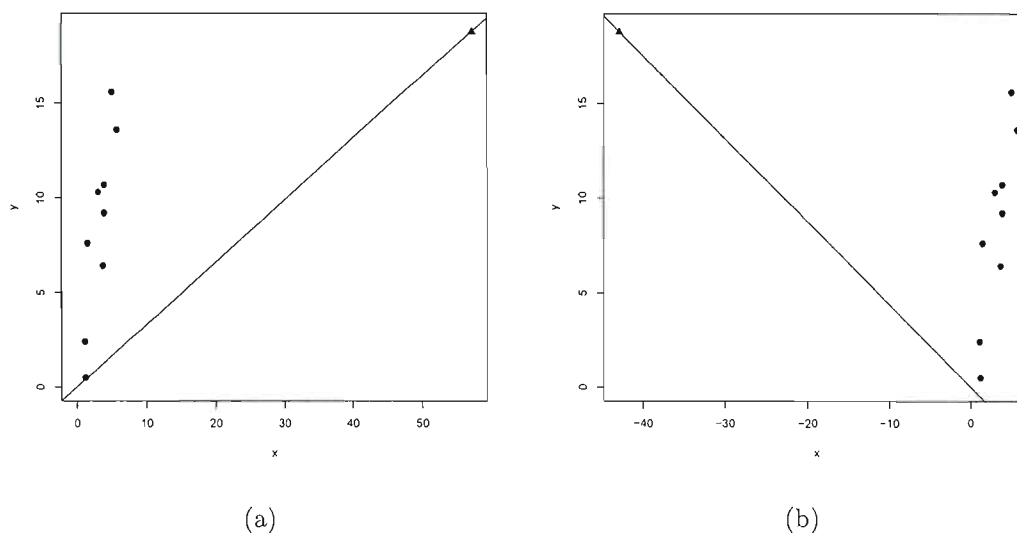


Figure 1.14 Estimé des moindres valeurs absolues de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,7; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

lorsque le point P_{10} est retiré de l'échantillon $(\hat{\sigma}_{L_1})$. L'estimation qui en résulte ne reflète pas l'allure de nos données. Le point aberrant, en exerçant une influence considérable sur la pente de la droite, affecte également l'estimation de σ . L'estimateur $\hat{\sigma}$ n'est pas robuste à une valeur aberrante en x .

Estimateur de Huber

L'estimateur de Huber a comme fonction objectif

$$\rho(u) = \begin{cases} u^2, & \text{si } |u| \leq m; \\ 2m|u| - m^2, & \text{si } |u| > m. \end{cases} \quad (1.11)$$

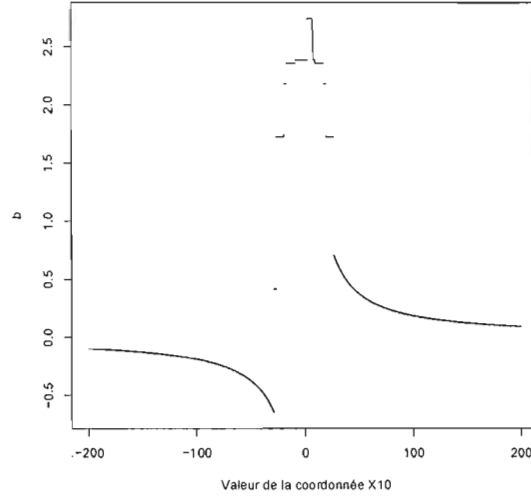


Figure 1.15 Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} .

La dérivée de $\rho(u)$ donne la fonction $\psi(u)$

$$\psi(u) = \begin{cases} 2u, & \text{si } |u| \leq m; \\ 2m \operatorname{sgn}(u), & \text{si } |u| > m. \end{cases} \quad (1.12)$$

où m est une constante. Enfin, la fonction $w(u)$ est donnée par

$$w(u) = \begin{cases} 2, & \text{si } |u| \leq m; \\ \frac{2m}{|u|}, & \text{si } |u| > m. \end{cases}$$

L'estimateur-M de Huber est un hybride entre l'estimateur des moindres carrés et l'estimateur des moindres valeurs absolues. Huber a considéré que l'estimateur des moindres carrés performe très bien lorsqu'il n'y a pas de valeurs aberrantes alors que l'estimateur des moindres valeurs absolues est très peu influencé par des valeurs y_i aberrantes. La fonction objectif d'un estimateur-M de Huber est donc quadratique en son centre et est

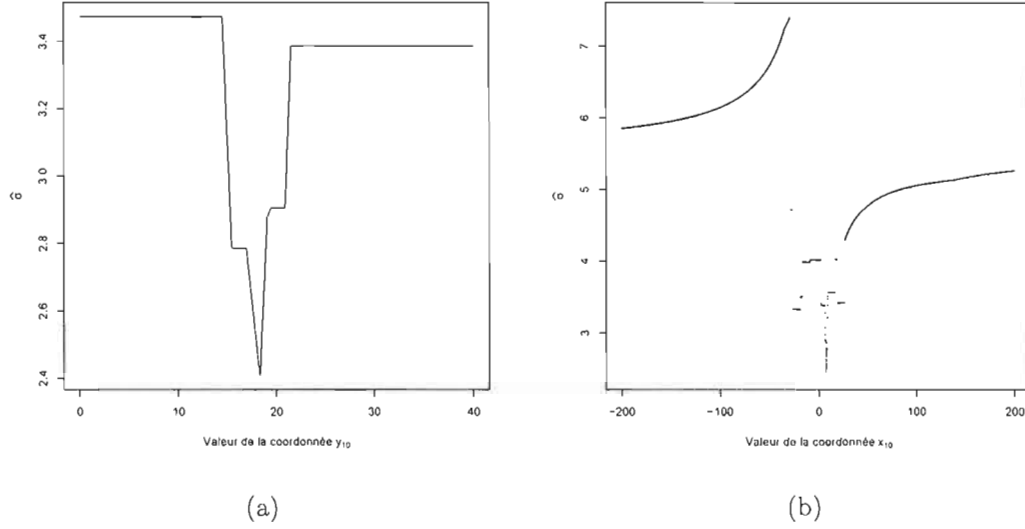


Figure 1.16 Comportement de l'estimé $\hat{\sigma}$ associé aux moindres valeurs absolues (a) pour différentes valeurs de la coordonnée y_{10} , (b) pour différentes valeurs de la coordonnée x_{10} .

linéaire dans les extrêmes. Nous allons maintenant analyser le comportement des estimateurs de β et σ en présence d'une valeur aberrante dans l'échantillon. Tel que suggérés par Birkes et Dodge (1993), nous avons fixé la constante à $m = 1,5$. Rappelons que $\frac{\varepsilon_i}{\sigma}$ i.i.d F , où F est non spécifiée. Toutefois, il est implicite que F a une dispersion semblable à une $N(0,1)$. On sait que $\Pr[-1,5 < Z < 1,5] = 0,866$ et que $\Pr[|Z| > 1,5] = 0,134$ si $Z \sim N(0,1)$. Par ce choix arbitraire de $m = 1,5$, la majorité des résidus seront traités selon les moindres carrés et une plus petite proportion, tout de même non négligeable de résidus à plus grande amplitude, sera traitée selon les moindres valeurs absolues. Les graphiques des fonctions $\rho(u)$, $\psi(u)$ et $w(u)$ sont illustrés sur les figures 1.17 (a), 1.17 (b) et 1.18, respectivement, pour $m = 1,5$. On a utilisé l'estimateur des MC comme valeur initiale $b^{(0)}$. Les poids $w_i^{(k)}$ du processus d'itération des moindres carrés repondérés sont donnés par

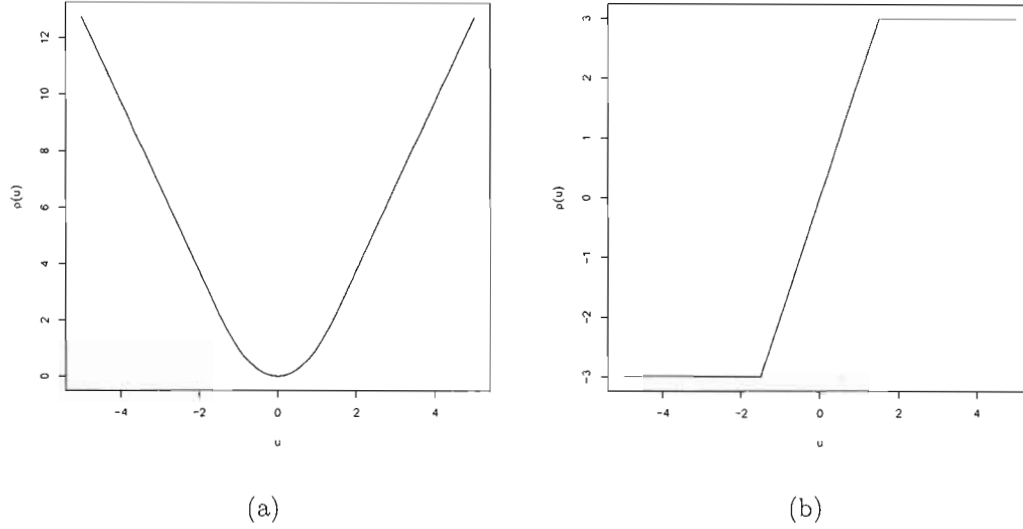


Figure 1.17 (a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés pour l'estimateur de Huber avec $m = 1,5$.

$$w_i^{(k)} = \begin{cases} 2, & \text{si } \frac{|e_i^{(k)}|}{\hat{\sigma}^{(k)}} \leq m ; \\ \frac{2m\hat{\sigma}^{(k)}}{|e_i^{(k)}|}, & \text{si } \frac{|e_i^{(k)}|}{\hat{\sigma}^{(k)}} > m. \end{cases} \quad (1.13)$$

Le cas où le point P_{10} est clairement aberrant en ordonnée est présenté sur les figures 1.19 (a) et (b). Les estimés de la pente sont $b_I = 2,9274$ et $b_{II} = 2,2555$ respectivement, en fixant $m = 1,5$. Ces valeurs sont peu éloignées de $b_H = 2,6178$, l'estimé Huber de β lorsque le point P_{10} est retiré de l'ensemble de données. La valeur aberrante a donc une faible influence sur l'estimation de β . Lorsque le point P_{10} est aberrant, l'estimateur de Huber se comporte, à une constante additive près, comme l'estimateur des moindres carrés calculé sur les points P_1, \dots, P_9 , puisque $\max\{|e_1|, \dots, |e_9|\} < m\hat{\sigma}$. Seul $|e_{10}| > m\hat{\sigma}$ a un poids w_{10} qui est inférieur à 2, et qui est d'autant plus petit que e_{10} est grand,

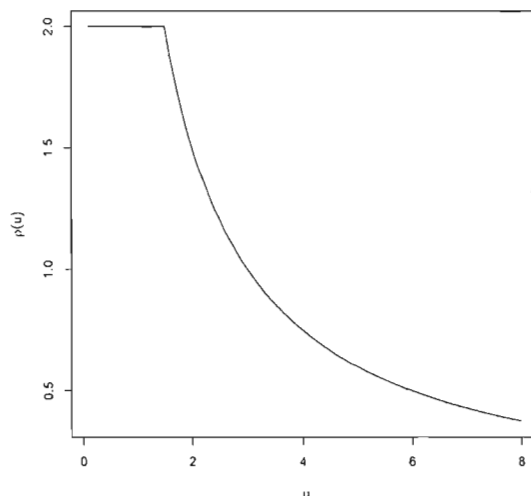


Figure 1.18 Fonction de poids reliée à la méthode itérative des moindres carrés pondérés pour l'estimateur de Huber avec $m = 1,5$.

puisque la droite passe dans la région des neuf points. On peut montrer que

$$\lim_{|y_{10}| \rightarrow \infty} w_{10} = \lim_{|y_{10}| \rightarrow \infty} \frac{2m \hat{\sigma}_I}{|e_{10}|} = 0,$$

où $\hat{\sigma}_I$ est l'estimé du paramètre σ lorsque y_{10} tend vers l'infini (nous verrons plus loin dans cette section que cet estimé converge lorsque y_{10} tend vers l'infini). L'estimateur de Huber, lorsque la coordonnée y_{10} tend vers plus l'infini, est alors calculé de la façon

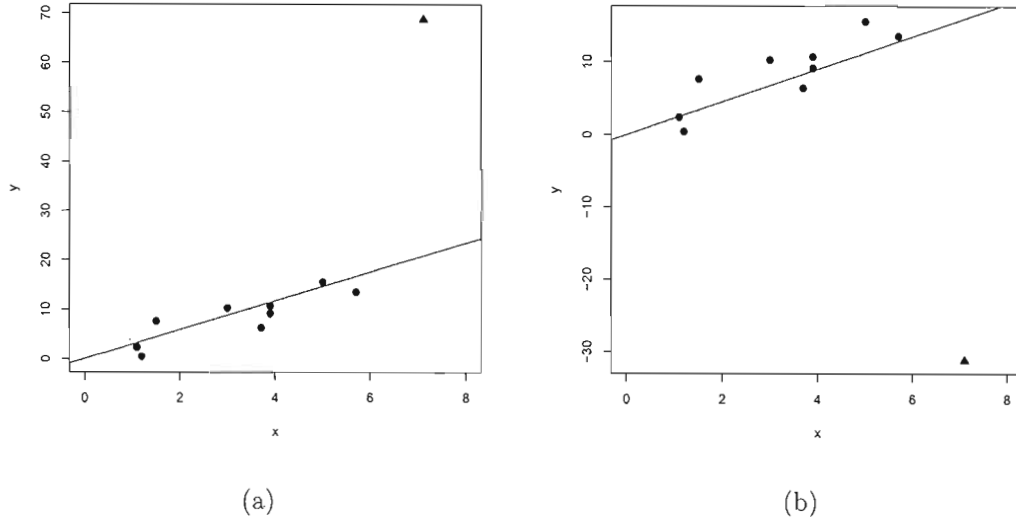


Figure 1.19 Estimé de Huber de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7, 1; 68, 8)$, (b) lorsque le point $P_{10}^{(II)} = (7, 1; -31, 2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

suivante :

$$\begin{aligned}
 b_I &= \lim_{y_{10} \rightarrow \infty} \frac{\sum_{i=1}^{10} x_i y_i w_i}{\sum_{i=1}^{10} x_i^2 w_i} \\
 &= \lim_{y_{10} \rightarrow \infty} \left[\frac{\sum_{i=1}^9 x_i y_i 2}{\sum_{i=1}^9 x_i^2 2} + \frac{x_{10} y_{10} w_{10}}{\sum_{i=1}^9 x_i^2 2} \right] \\
 &= \frac{604,7}{231} + \frac{x_{10}}{2 \sum_{i=1}^9 x_i^2} \lim_{y_{10} \rightarrow \infty} y_{10} w_{10} \\
 &= 2,6178 + \frac{x_{10}}{2 \sum_{i=1}^9 x_i^2} \lim_{y_{10} \rightarrow \infty} y_{10} \frac{2m \hat{\sigma}_I}{|y_{10} - b_I x_{10}|} \\
 &= b_{MC} + \frac{x_{10}}{2 \sum_{i=1}^9 x_i^2} 2m \hat{\sigma}_I
 \end{aligned}$$

où b_{MC} est l'estimé de β des moindres carrés obtenu à la section 1.2 lorsque le point P_{10} est retiré de l'ensemble de données. Notons que $b_{MC} = b_H$. Ainsi, $b_I = b_{MC} + 0,3096 = 2,9274$. De façon similaire, lorsque y_{10} tend vers moins l'infini, on obtient

$$b_{II} = b_{MC} - \frac{x_{10}}{2 \sum_{i=1}^9 x_i^2} 2m \hat{\sigma}_{II},$$

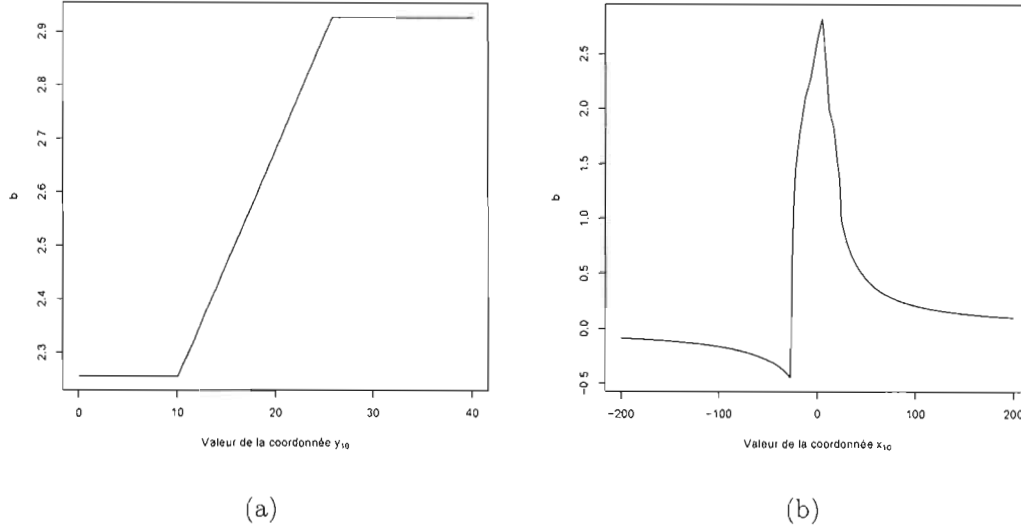


Figure 1.20 Comportement de l'estimé b de Huber pour (a) différentes valeurs de la coordonnée y_{10} (b) différentes valeurs de la coordonnée x_{10} .

où $\hat{\sigma}_{II}$ est l'estimé du paramètre σ lorsque y_{10} tend vers moins l'infini. Donc, $b_{II} = b_{MC} - 0,3623 = 2,2555$. Lorsque $y_{10} \in (10,2; 25,8)$, correspondant au segment de droite dont la pente est non nulle sur la figure 1.20 (a), aucune valeur résiduelle n'est plus grande que la constante $m = 1,5$. Chaque point se voit alors attribuer le même poids et l'estimateur de Huber est en fait l'estimateur des MC appliqué sur les dix points. L'estimateur de Huber est robuste en présence d'une valeur aberrante en ordonné, mais comme l'estimateur des moindres valeurs absolues, il manque de finesse. On voit sur la figure 1.20 (a) que l'estimateur de Huber, qui surestime b_H , est égal à 2,9274 peu importe que l'ordonnée du point P_{10} soit égale à 30 ou à 100 000, par exemple.

Considérons maintenant le cas où le point P_{10} est clairement aberrant en abscisse. Sur les graphiques (a) et (b) de la figure 1.21, on observe les droites dont les pentes sont estimées à $b_{III} = 0,3859$ et $b_{IV} = -0,3231$ respectivement. On remarque très bien sur ces graphiques que les points $P_{10}^{(III)} = (57,1; 18,8)$ et $P_{10}^{(IV)} = (-42,9; 18,8)$ respectivement, ont une emprise importante sur la droite de telle sorte que celle-ci dévie de la tendance

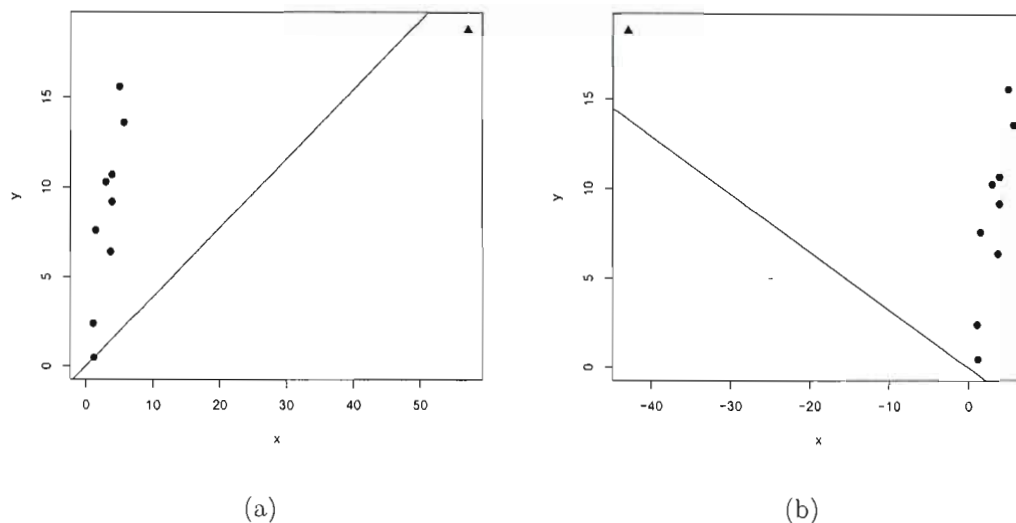


Figure 1.21 Estimé de Huber de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,7; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

linéaire des 9 autres points en empruntant une trajectoire dans laquelle ne se trouve presque aucun point. On observe sur la figure 1.20 (b) que plus l'abscisse du point P_{10} tend vers plus ou moins l'infini, plus la pente de la droite tend vers 0. Les résidus étant la distance verticale entre les points et la droite, cette dernière aura tendance à se rapprocher de P_{10} pour éviter d'énormes résidus. Lorsque la convergence est atteinte, disons à l'itération k , pour les dix points de l'ensemble, $\frac{|e_i|}{\hat{\sigma}}$ est soit plus petit que $m = 1,5$, soit juste un peu plus grand de telle sorte que les poids w_i sont à peu près les mêmes pour les dix points, soit de 2 où légèrement supérieurs. L'estimateur de Huber se comporte donc à peu près comme l'estimateur des MC. La méthode n'est pas robuste à une valeur aberrante en x . On notera que pour certaines valeurs de x_{10} , on a observé que l'estimé b oscille entre deux valeurs. Par exemple, lorsqu'on considère $x_{10} = -29,3$, on obtient les valeurs $-0,4199$ et $-0,4335$. Hoaglin et al. (1985) mentionnent qu'il est possible que l'équation (1.8) avec la fonction ψ de Huber donne plus d'une solutions.

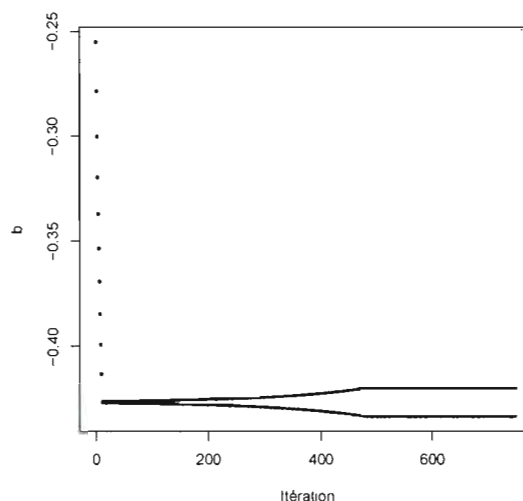


Figure 1.22 L'estimé de b en fonction du nombre d'itérations effectuées avec la méthode des moindres carrés repondérés, lorsque $x_{10} = -29,3$.

La figure 1.22 illustre le comportement de l'estimé b en fonction du nombre d'itérations pour $x_{10} = -29,3$.

On étudie maintenant le comportement de l'estimateur de σ en présence d'une valeur aberrante. Lorsque P_{10} est aberrant en ordonnée, l'influence qu'il détient sur l'estimation de σ est limitée. En effet, on observe sur le graphique (a) de la figure 1.23 que lorsque le point P_{10} tend vers plus ou moins l'infini, l'estimé de σ cesse de croître et devient constant à 3,3586 ou 3,9281 respectivement. Il s'agit de valeurs qui surestiment $\hat{\sigma}_H = 3,2051$, soit l'estimé de σ calculé en omettant le point P_{10} . Le comportement de $\hat{\sigma}$, lorsque y_{10} tend vers plus ou moins l'infini, s'explique de façon similaire au comportement de $\hat{\sigma}$ relié à l'estimateur L_1 de β vu à la section précédente. Il s'agit donc d'un estimateur robuste à une valeur aberrante en y , mais qui manque de finesse en produisant la même valeur de $\hat{\sigma}$ pour un point dont l'ordonnée vaut 30 que pour un point dont $x_{10} = 100\,000$, par exemple.

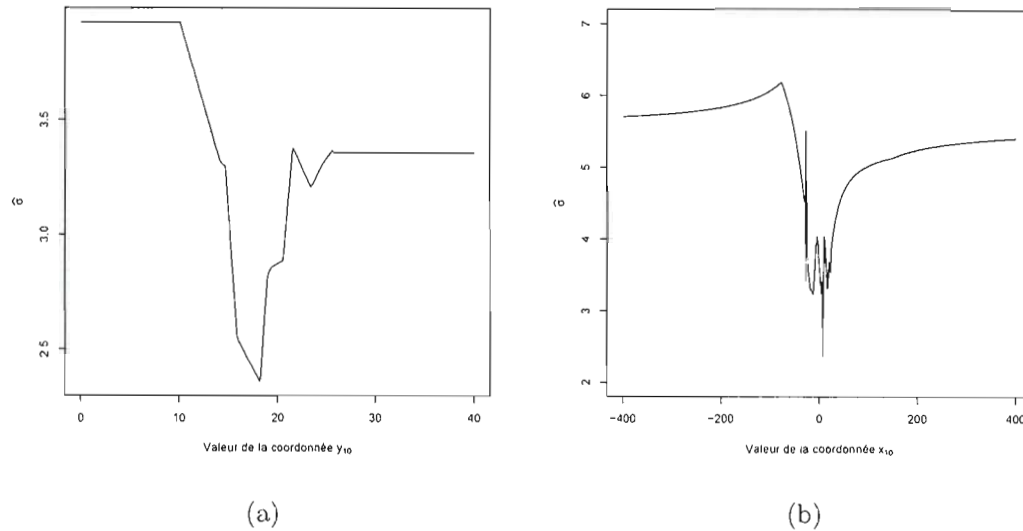


Figure 1.23 Comportement de l'estimé $\hat{\sigma}$ associé à Huber (a) pour différentes valeurs de la coordonnées y_{10} , (b) pour différentes valeurs de la coordonnée x_{10} .

Le graphique (b) de la figure 1.23 illustre le comportement de $\hat{\sigma}$ lorsque le point P_{10} est aberrant en abscisse. On peut y voir que $\hat{\sigma}$ converge vers la valeur 5,5598 lorsque x_{10} tend vers plus ou moins l'infini. Cependant, comme pour l'estimé du paramètre d'échelle associé à l'estimateur L_1 de β , les graphiques (a) et (b) de la figure 1.21 indiquent que la majorité des résidus ont des valeurs plus grandes que $\hat{\sigma}_H$. Ces grands résidus sont à l'origine de la surévaluation du paramètre d'échelle. Le point aberrant P_{10} , en influençant l'estimateur b de Huber affecte l'estimation de σ , qui s'avère non robuste par rapport à une valeur aberrante en x .

Estimateur d'Andrews

L'estimateur d'Andrews, aussi appelé *Andrew's waves* ou estimateur sinus, a pour fonction objectif

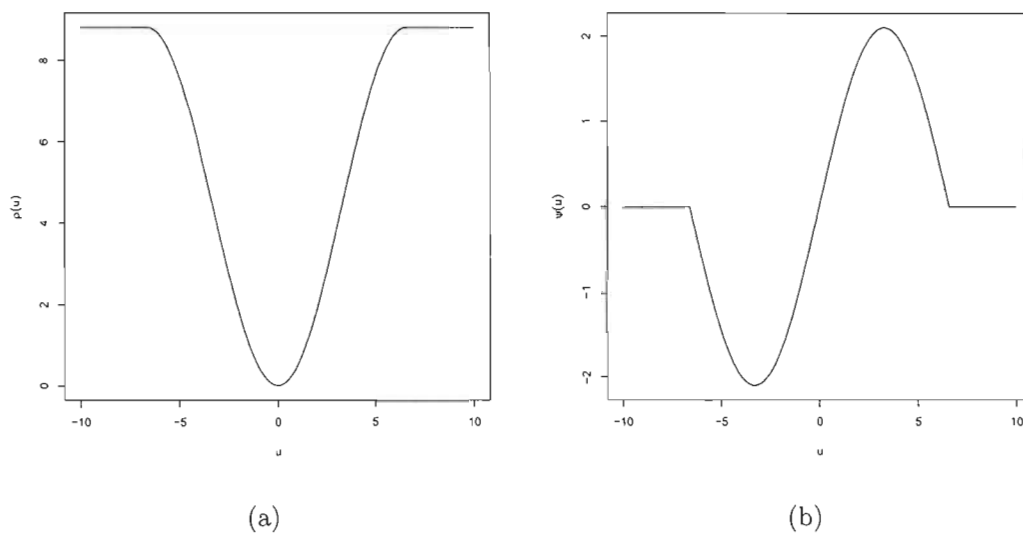


Figure 1.24 (a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés pour l'estimateur d'Andrews avec constante $A = 2,1$.

$$\rho(u) = \begin{cases} A^2 \left[1 - \cos\left(\frac{u}{A}\right) \right], & \text{si } |u| \leq \pi A; \\ 2A^2, & \text{si } |u| > \pi A. \end{cases}$$

Sa fonction ψ , est donnée par

$$\psi(u) = \begin{cases} A \sin\left(\frac{u}{A}\right), & \text{si } |u| \leq \pi A; \\ 0, & \text{si } |u| > \pi A. \end{cases}$$

Les fonctions $\rho(u)$ et $\psi(u)$ sont illustrées sur les figures 1.24 (a) et 1.24 (b), respectivement, pour $A = 2,1$. L'estimateur d'Andrews appartient à la classe d'estimateurs-M redescendants (*redescending M-estimators*, c'est-à-dire $\psi(u) = 0$ pour un certain nombre positif t tel que $|u| > t$, voir Hampel et al. (1986), Lawrence et Arthur (1990)). Pour

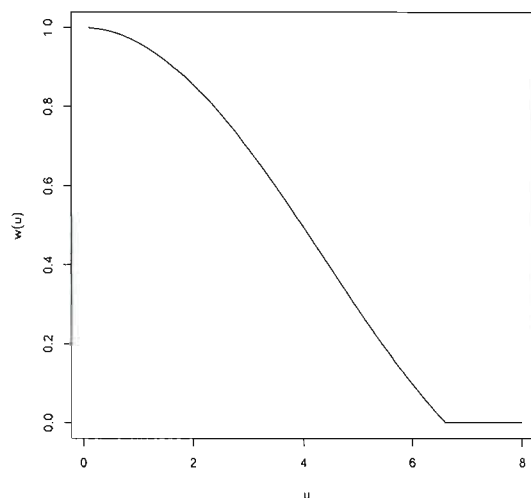


Figure 1.25 Fonction de poids reliée à la méthode itérative des moindres carrés repondérés pour l'estimateur d'Andrews avec constante $A = 2,1$.

cette classe d'estimateurs, Hoalgin et al. (1985) suggèrent de choisir une valeur initiale de β en procédant de la façon suivante. Tout d'abord, on choisit un estimateur-M dont la fonction ψ est monotone, comme l'estimateur L_1 . On utilise le processus d'itération jusqu'à ce qu'on ait atteint une précision suffisante pour calculer l'estimé des moindres valeurs absolues. Enfin, pour calculer l'estimateur d'Andrews, on utilise comme valeur initiale b_0 , l'estimé obtenu à l'étape d'avant. Plusieurs valeurs de A ont été proposées par différents auteurs (voir Hoalgin et al., 1983). Entre autres, Hogg utilise des valeurs de A égales à 1,5 et 2, Gross suggère de prendre $A = 1,8$ ou $A = 2,4$, alors que Andrews et al. (1972) proposent de prendre $A = 2,1$. De façon arbitraire, nous avons choisi $A = 2,1$. Le processus itératif des moindres carrés repondérés accorde à chaque point le poids suivant :

$$w_i^{(k)} = \begin{cases} \frac{\sigma A \sin\left(\frac{e_i^{(k)}}{\hat{\sigma}^{(k)} A}\right)}{e_i^{(k)}}, & \text{si } \left|\frac{e_i^{(k)}}{\hat{\sigma}^{(k)}}\right| \leq \pi A; \\ 0, & \text{si } \left|\frac{e_i^{(k)}}{\hat{\sigma}^{(k)}}\right| > \pi A. \end{cases} \quad (1.14)$$

La figure 1.25 présente la fonction de poids $w(u)$ avec $A = 2,1$.

On considère d'abord le cas où le point P_{10} est clairement aberrant en ordonnée en utilisant les points $P_{10}^{(I)} = (7,1;68,8)$ et $P_{10}^{(II)} = (7,1;-31,2)$. Les graphiques (a) et (b) de la figure 1.26 illustrent que la droite suit la tendance linéaire des points P_i pour $i = 1, \dots, 9$. Les estimés pour les 2 graphiques sont $b_I = 2,6171$ et $b_{II} = 2,6169$ respectivement. La droite ne se laisse pas influencer par les points $P_{10}^{(I)}$ et $P_{10}^{(II)}$ respectivement, compte tenu du fait que l'estimateur d'Andrews, lorsque le point P_{10} n'exerce aucune influence sur la droite, est $b_A = 2,6168$. On voit par l'équation (1.14) que le processus itératif des moindres carrés repondérés n'accorde aucun poids ($w_i = 0$) à un point dont la valeur résiduelle est trop grande, ce qui est exactement le cas lorsque y_{10} tend vers plus ou moins l'infini. L'estimateur de β est alors calculé en tenant compte uniquement des points P_1, P_2, \dots, P_9 fixés. L'estimé b devient alors constant et égal à b_I et b_{II} respectivement, tel qu'illustré sur la figure 1.27 (a). La technique d'estimation rejette le point aberrant P_{10} . L'estimateur b d'Andrews est clairement robuste à une valeur aberrante en y . La légère différence observée entre b_I et b_{II} provient du fait que les poids $w_i^{(k)}$ accordés à chaque point diffèrent lorsque y_{10} tend vers soit plus l'infini, soit moins l'infini. En analysant les données, nous avons constaté que l'estimé du paramètre d'échelle tend vers deux valeurs différentes correspondant aux cas où y_{10} tend vers plus ou moins l'infini (on remarque cette situation sur le graphique (b) de la figure 1.27. Nous y reviendrons un peu plus loin). Les poids $w_i^{(k)}$ de l'équation (1.14) dépendent de l'estimé du paramètre d'échelle, d'où la différence entre b_I et b_{II} .

Le graphique (b) de la figure 1.27 dépeint le comportement de l'estimé $\hat{\sigma}$ lorsque l'ordonnée du point P_{10} varie. On remarque que lorsque la coordonnée y_{10} tend vers plus ou moins l'infini, l'estimé $\hat{\sigma}$ devient constant et vaut $\hat{\sigma}_I = 3,6686$ ou $\hat{\sigma}_{II} = 3,2901$

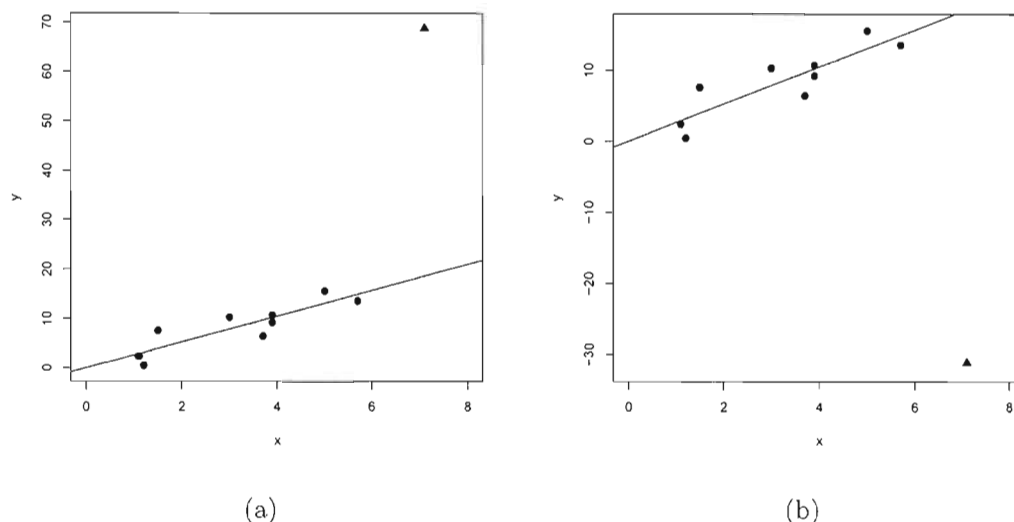


Figure 1.26 Estimé d'Andrews de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7, 1; 68, 8)$, (b) lorsque le point $P_{10}^{(II)} = (7, 1; -31, 2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

respectivement. Tout comme les estimateurs-M vu précédemment, le fait que $\hat{\sigma}$ défini par l'équation (1.9) devient constant est dû à la résistance de la médiane par rapport à une valeur aberrante et au fait que l'estimateur b d'Andrews devient lui aussi constant lorsque l'ordonnée du point P_{10} tend vers l'infini (voir graphique (a) de la figure 1.27). On constate que $\hat{\sigma}_I$ et $\hat{\sigma}_{II}$ s'apparentent à $\hat{\sigma}_A = 3,2049$, c'est-à-dire, l'estimateur de σ lorsque le point P_{10} est retiré des données. On conclut que $\hat{\sigma}$ associé à l'estimateur-M d'Andrews est un estimateur robuste à une valeur aberrante en y qui manque de finesse, tout comme les estimateurs-M vus précédemment.

On considère maintenant le cas où le point P_{10} est clairement aberrant en abscisse. Les points $P_{10}^{(III)} = (57, 7; 18, 8)$ et $P_{10}^{(IV)} = (-42, 9; 18, 8)$ des graphiques (a) et (b) de la figure 1.28, attirent la droite, forçant cette dernière à emprunter une trajectoire dans laquelle ne se trouve presque aucun point. Les estimés sont $b_{III} = 0,3909$ et $b_{IV} = -0,3029$ respectivement. Ces estimés sont non conformes à la tendance linéaire

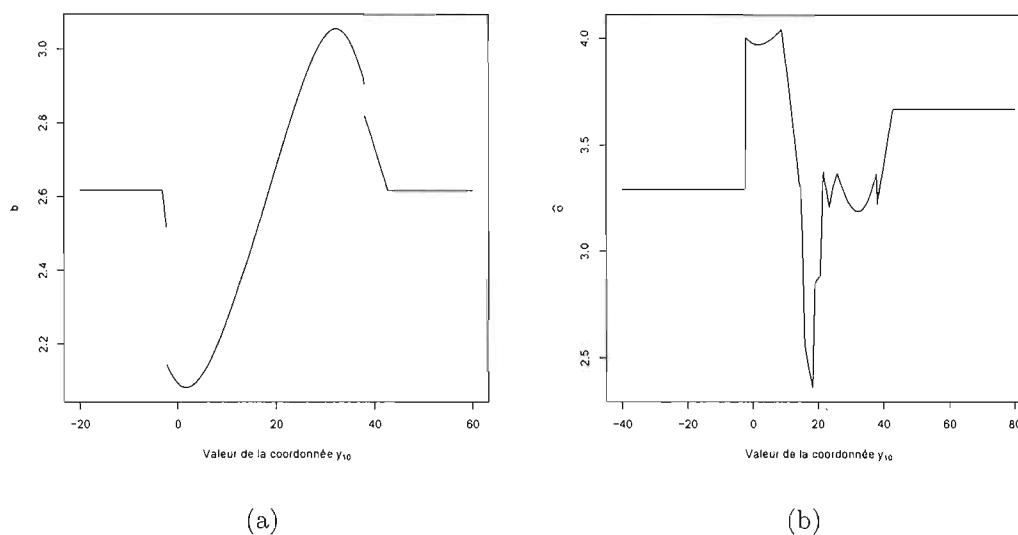


Figure 1.27 Comportement, pour différentes valeurs de la coordonnées y_{10} , (a) de l'estimé b et (b) de l'estimé $\hat{\sigma}$, tous deux associés à Andrews.

décrite par les neuf points non aberrants. Le graphique (a) de la figure 1.29 décrit le comportement de b , à savoir, lorsque x_{10} tend vers plus ou moins l'infini, l'estimateur de β tend vers 0. L'explication est similaire à celle donnée pour l'estimateur-M de Huber. Lorsque la convergence est atteinte à l'itération k , les quantités $\frac{e_i^{(k)}}{\hat{\sigma}^{(k)}}$ sont inférieures à πA pour tout $i = 1, \dots, 10$. De plus, les poids $w_i^{(k)}$, pour $i = 1, \dots, 10$, sont tous à peu près égaux. L'estimateur-M d'Andrews se comporte de façon similaire à l'estimateur des MC. L'estimé b d'Andrews n'est pas robuste à une valeur aberrante en x . Tout comme pour l'estimateur-M de Huber, on observe que pour certaines valeurs de x_{10} , la méthode itérative des moindres carrés repondérés produit une solution qui oscille entre deux valeurs. Hoaglin et al. (1985) discutent brièvement de telles situations lorsqu'on travaille avec un estimateur-M redescendant. Ils mentionnent que pour ce type d'estimateur, la solution de l'équation (1.8) peut ne pas être unique, ou encore, le procédé itératif pourrait mener à des minimums locaux.

Sur le graphique (b) de la figure 1.29, on peut observer le comportement de l'estimateur

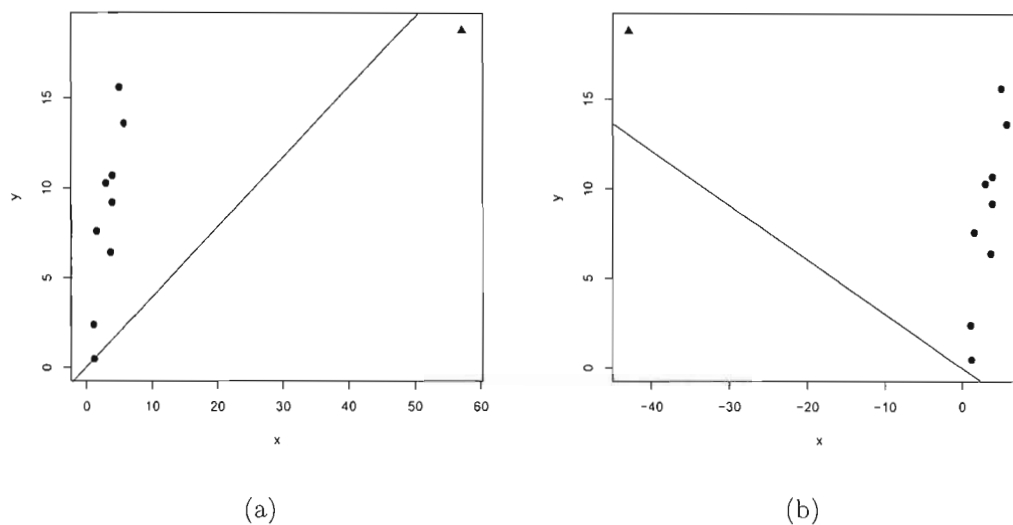


Figure 1.28 Estimé d'Andrews de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,7; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

de σ lorsque le point P_{10} tend vers plus ou moins l'infini en abscisse. Tout comme l'estimateur de Huber, lorsque x_{10} tend vers plus ou moins l'infini, l'estimation de σ converge vers un estimé $\hat{\sigma} = 5,5598$, qui surestime $\hat{\sigma}_A = 3,2049$. Tout comme pour $\hat{\sigma}$ associé à l'estimateur-M de Huber, la majorité des résidus ont une amplitude plus importante que lorsque le point P_{10} est retiré de l'échantillon ($\hat{\sigma}_A$). L'influence que détient le point aberrant P_{10} sur l'estimation de β s'étend aussi à l'estimation du paramètre d'échelle de telle sorte que $\hat{\sigma}$ n'est pas robuste en présence d'une valeur aberrante en x .

Estimateur bipoids de Tukey

L'estimateur bipoids de Tukey (*Tukey's biweight*) aussi appelé estimateur bicarré (*bisquare estimator*) a comme fonction objectif

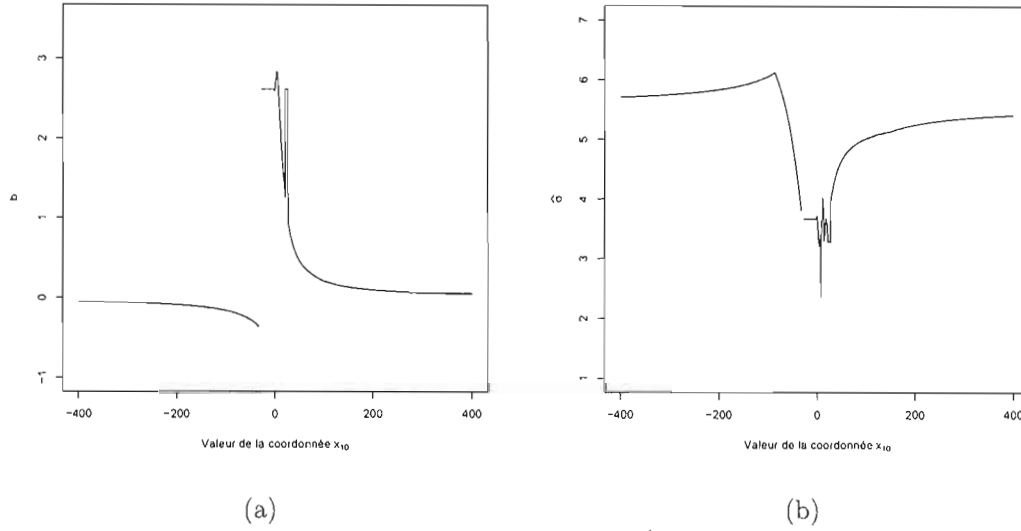


Figure 1.29 Comportement, pour différentes valeurs de la coordonnées x_{10} , (a) de l'estimé $\hat{\beta}$ et (b) de l'estimé $\hat{\sigma}$, tous deux associés à Andrews.

$$\rho(u) = \begin{cases} \frac{B^2}{6} \left\{ 1 - \left[1 - \left(\frac{u}{B} \right)^2 \right]^3 \right\}, & \text{si } |u| \leq B; \\ \frac{B^2}{6}, & \text{si } |u| > B. \end{cases}$$

Sa fonction ψ obtenue en dérivant ρ est

$$\psi(u) = \begin{cases} u \left[1 - \left(\frac{u}{B} \right)^2 \right]^2, & \text{si } |u| \leq B; \\ 0, & \text{si } |u| > B. \end{cases}$$

L'estimateur de Tukey, tout comme celui d'Andrews, est un estimateur-M redescendant. La valeur initiale de b a été choisie de la même façon que pour l'estimateur d'Andrews. Pour ce qui est de la valeur constante B , Hoaglin et al. (1983) proposent de la choisir telle que $6 \leq B \leq 12$. Le choix de B dans cet intervalle permet d'obtenir un estimateur-M

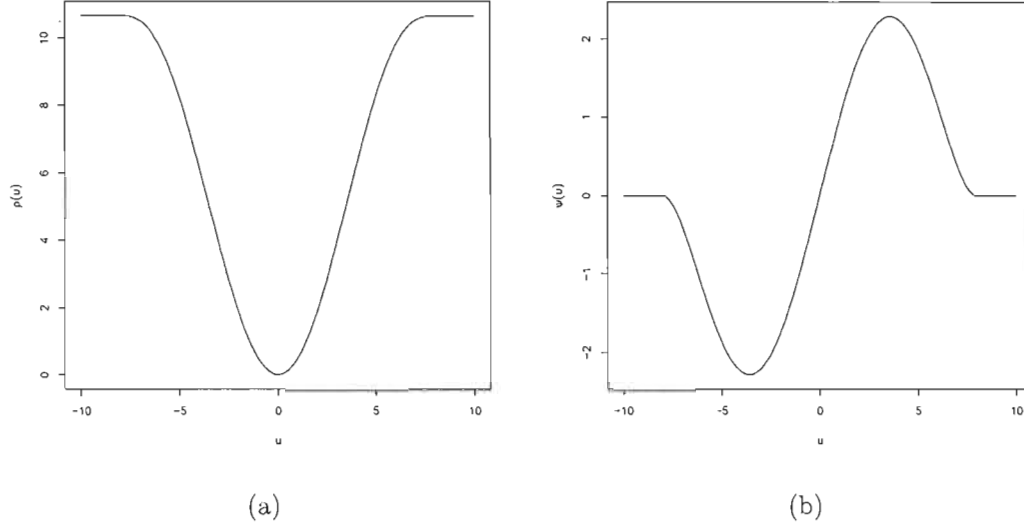


Figure 1.30 (a) Fonction objectif et (b) fonction ψ reliées à la méthode itérative des moindres carrés repondérés, pour l'estimateur bipoids de Tukey avec constante $B = 8$.

bipoids qui est résistant aux valeurs aberrantes. Nous avons choisi, de façon arbitraire, $B = 8$. Le processus d'itération des moindres carrés repondérés accorde un poids, associé à l'estimateur b de Tukey, de

$$w_i^{(k)} = \begin{cases} \left[1 - \left(\frac{e_i^{(k)}}{\hat{\sigma}^{(k)} B} \right)^2 \right]^2, & \text{si } \left| \frac{e_i^{(k)}}{\hat{\sigma}^{(k)}} \right| \leq B; \\ 0, & \text{si } \left| \frac{e_i^{(k)}}{\hat{\sigma}^{(k)}} \right| > B. \end{cases}$$

Les fonctions $\rho(u)$, $\psi(u)$ et $w(u)$ sont illustrées sur les figures 1.30 (a), 1.30 (b) et 1.31, respectivement, pour $B = 8$.

Considérons d'abord le cas où le point P_{10} est clairement aberrant en ordonnée. Ce cas est illustré sur les graphiques (a) et (b) de la figure 1.32. Malgré que les points $P_{10}^{(I)}$ et $P_{10}^{(II)}$, respectivement, s'écartent de la tendance linéaire formée par les points restants,

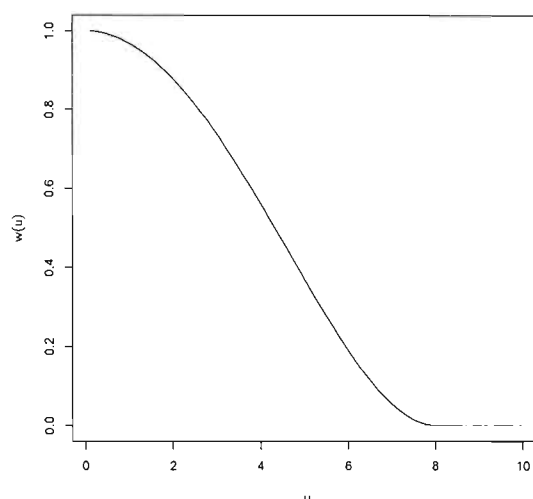


Figure 1.31 Fonction de poids reliée à la méthode itérative des moindres carrés repondérés, pour l'estimateur bipoids de Tukey avec constante $B = 8$.

la droite ne semble en aucun cas être affectée. Pour les deux graphiques, l'estimé de β est $b_I = b_{II} = 2,6178$ respectivement. Compte tenu du fait que l'estimé b de Tukey, lorsque le point P_{10} est enlevé des données, est $b_T = 2,6178$, on voit que la technique d'estimation écarte ce point. Elle n'en tient pas compte dans l'estimation de β . Lorsqu'il y a convergence à l'itération k , alors pour une valeur résiduelle trop éloignée de B , c'est-à-dire lorsque y_{10} tend vers plus ou moins l'infini, le processus itératif attribue à ce point un poids $w_{10}^{(k)}$ égal à zéro, tel qu'illustré sur la figure 1.33. Ce point n'a aucune influence sur la droite et est considéré par le modèle comme une valeur aberrante. Il est clair que l'estimateur-M de Tukey est robuste à une valeur aberrante en y .

On considère maintenant le cas où le point P_{10} est aberrant en abscisse. Les points $P_{10}^{(III)} = (57,1; 18,8)$ et $P_{10}^{(IV)} = (-42,9; 18,8)$ détiennent une grande influence sur l'estimation de β comme l'illustrent les graphiques (a) et (b) de la figure 1.34. La droite se compromet en suivant un trajet qui se retrouve quelque part entre le point aberrant et

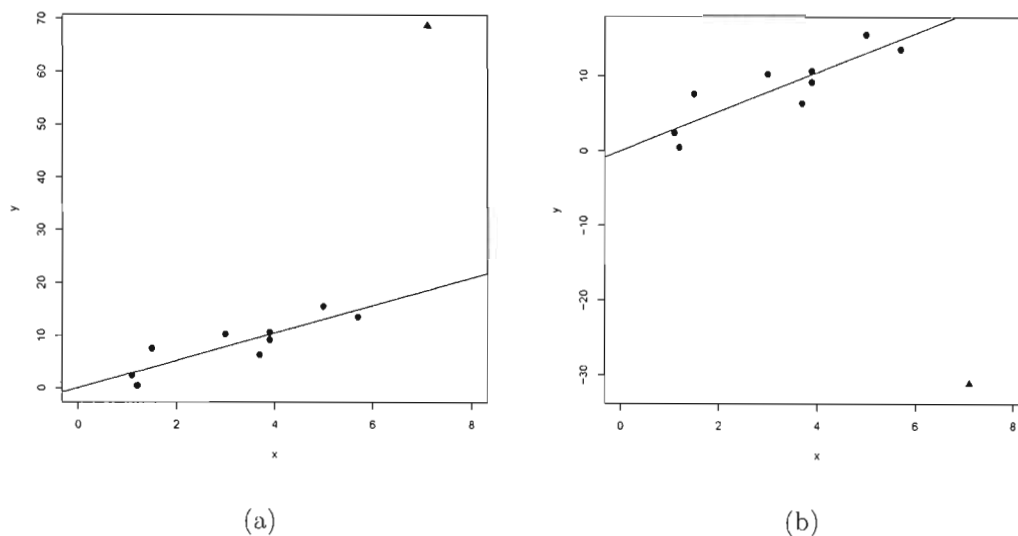


Figure 1.32 Estimé bipoids de Tukey de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7, 1; 68, 8)$, (b) lorsque le point $P_{10}^{(II)} = (7, 1; -31, 2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

le reste de l'ensemble. Les estimés de la pente lorsque le point P_{10} est aberrant, donnés par $b_{III} = 0,4075$ et $b_{IV} = -0,2578$ respectivement, ne sont pas représentatifs de l'estimé de β lorsque le point P_{10} n'a pas d'influence, c'est-à-dire $b_I = 2,6178$. Lorsque ce point tend vers plus ou moins l'infini, la pente de la droite converge vers la valeur 0, telle qu'illustrée sur la figure 1.35. L'explication du comportement de b est similaire à celui des estimateurs-M de Huber et Andrews. Lorsqu'il y a convergence dans le processus itératif des moindres carrés repondérés, l'estimateur-M de Tukey se comporte de façon similaire à l'estimateur des MC puisque les poids $w_i^{(k)}$ associés sont à peu près les mêmes. Cela explique que l'estimateur-M de Tukey n'est pas robuste à une valeur aberrante en x .

La robustesse de l'estimateur $\hat{\sigma}$ en présence d'une valeur aberrante tant en ordonnée qu'en abscisse, est analysée grâce, entre autres, aux graphiques (a) et (b), respectivement de la figure 1.36. L'estimé de σ lorsque le point P_{10} est retiré de l'ensemble est donné

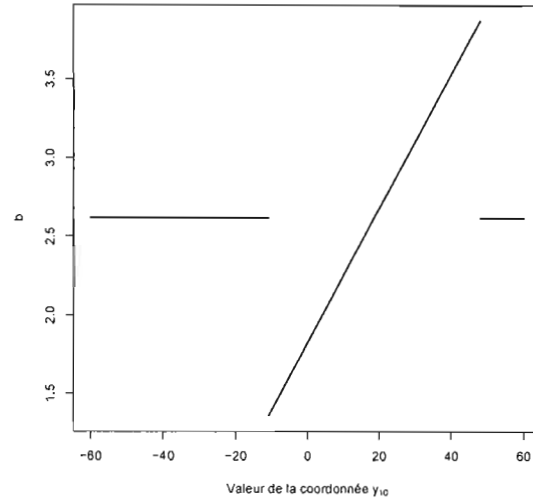


Figure 1.33 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} .

par $\hat{\sigma}_T = 3,2051$. On voit sur le graphique (a) de la figure 1.36 que lorsque le point P_{10} tend vers plus ou moins l'infini, l'estimé de σ devient constant et prend les valeurs 3,667 et 3,29 respectivement. L'explication du comportement de $\hat{\sigma}$ lorsque l'ordonnée du point P_{10} tend vers plus ou moins l'infini est semblable à celle du comportement de $\hat{\sigma}$ associé à l'estimateur d'Andrews. Cela découle de la similarité de leur fonction ψ comme en témoigne les graphiques des figures 1.24 et 1.30. L'emploi de la médiane dans le calcul de $\hat{\sigma}$ et le fait que l'estimateur de β soit constant pour de grandes valeurs de y_{10} , tel qu'illustré sur la figure 1.33, font en sorte que l'estimé du paramètre d'échelle soit constant égal à 3,667 ou 3,29 lorsque l'ordonnée du point P_{10} tend vers plus ou moins l'infini respectivement. L'estimateur de σ associé à l'estimateur-M de Tukey est robuste à une valeur aberrante en y , mais manque de finesse tout comme ce fut le cas pour les estimateurs-M vus précédemment. Lorsque le point P_{10} est aberrant en abscisse, l'estimateur de σ n'est pas robuste. On observe sur le graphique (b) de la figure 1.36 que l'influence que détient le point aberrant sur $\hat{\sigma}$ est limité, car ce dernier converge vers $\hat{\sigma} = 5,559$ qui surévalue cependant $\hat{\sigma}$. À nouveau, les valeurs résiduelles plus grandes que

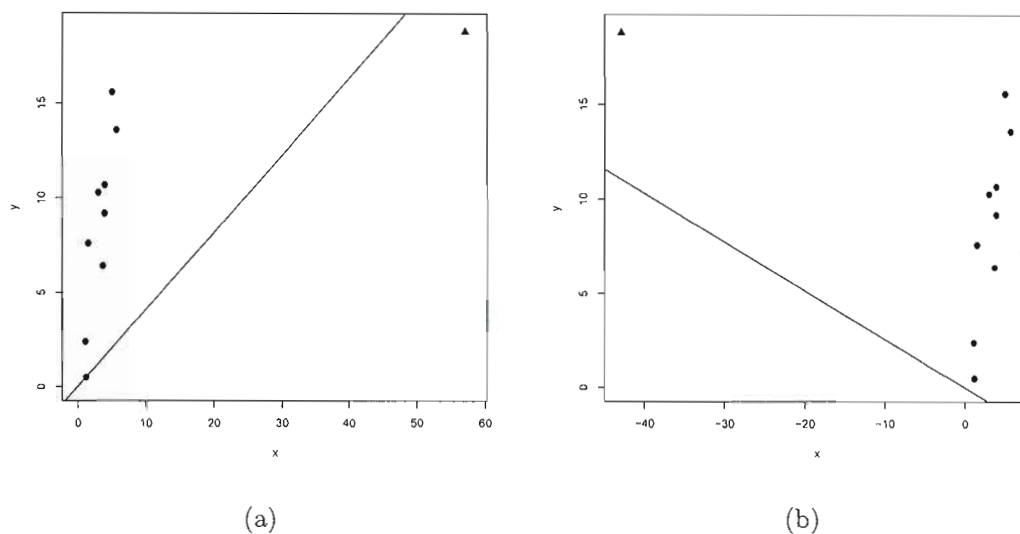


Figure 1.34 Estimé bipoids de Tukey de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,7; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

$\hat{\sigma}_T$ étant majoritaire, les médianes dans le calcul de l'estimateur du paramètre d'échelle en seront affectées, d'où la surestimation du paramètre σ .

1.4 Droite résistante de Tukey

La droite résistante (*resistant line*) a été proposée par Tukey (voir Johnstone et Velleman, 1985) afin de permettre d'ajuster une droite de la forme $y_i = \alpha + \beta x_i$ à un ensemble de points $(x; y)$. Nous présenterons donc dans cette section la méthode permettant d'estimer les paramètres du modèle

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

où $i = 1, \dots, n$. L'emphasis sera toutefois mis sur le paramètre d'intérêt de ce mémoire, soit β . La méthode ne permet pas d'estimer la pente β du modèle étudié dans ce mémoire, $y_i = \beta x_i + \varepsilon_i$. Puisque la tendance décrite par les données du tableau 1.1

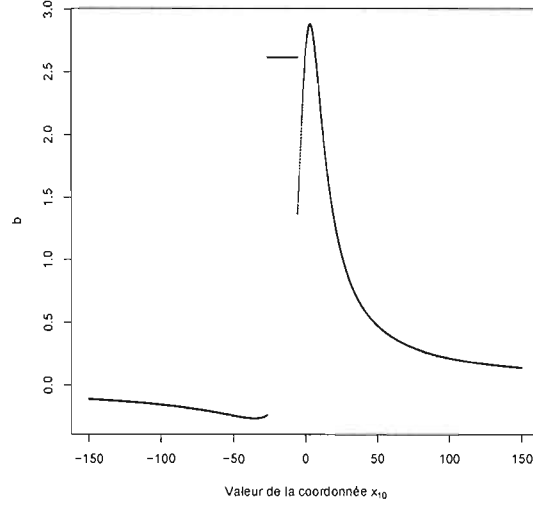


Figure 1.35 Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} .

suggère que la vraie droite pourrait passer par l'origine, nous allons simplement poser $\alpha = 0$.

Soit un ensemble de n points de la forme $(x_i, y_i), i = 1, \dots, n$ et soit les statistiques d'ordre $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. On divise l'ensemble de données en trois sous-ensembles, notés G , C , et D de taille respective n_G, n_C et n_D , avec $n_G + n_C + n_D = n$, et tel que

$$G = \{(x_i, y_i) \mid x_i \leq x_{(n_G)}\} \quad (1.15)$$

$$C = \{(x_i, y_i) \mid x_{(n_G+1)} \leq x_i \leq x_{(n-n_D)}\} \quad (1.16)$$

$$D = \{(x_i, y_i) \mid x_{(n-n_D+1)} \leq x_i \leq x_{(n)}\}. \quad (1.17)$$

Alors a et b sont les estimateurs de α et β , si a et b sont solution en α et β de l'équation

$$\text{med}_{(x_i, y_i) \in G} (y_i - \alpha - \beta x_i) = \text{med}_{(x_i, y_i) \in D} (y_i - \alpha - \beta x_i), \quad (1.18)$$

où $\text{med}_{(x_i, y_i) \in G} (y_i - \alpha - \beta x_i)$ est la médiane de l'ensemble des erreurs $\varepsilon_i = y_i - \alpha - \beta x_i$ associées à tous les couples $(x_i, y_i) \in G$. Lorsque l'échantillon de données le permet, on divise les données de façon à ce que $n_G = n_D$. Cependant, s'il y a des égalités

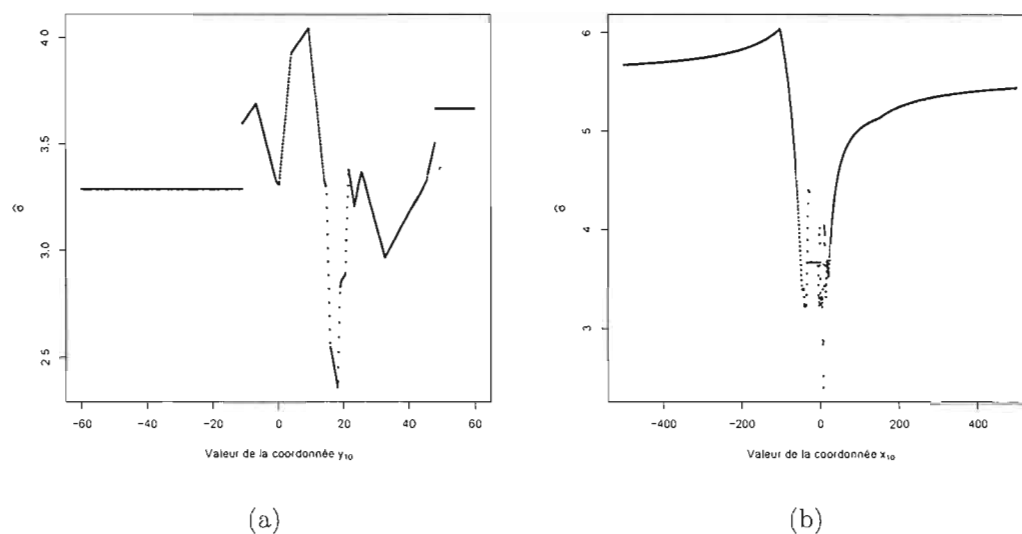


Figure 1.36 Comportement de l'estimé $\hat{\sigma}$ associé à Tukey (a) pour différentes valeurs de la coordonnée y_{10} , (b) pour différentes valeurs de la coordonnée x_{10} .

entre certaines valeurs x_i , on distribue les valeurs égales dans le même groupe. On doit cependant s'assurer que les sous-ensembles G et D contiennent au moins trois points chacun afin d'obtenir une certaine robustesse par rapport à une valeur aberrante. Un grand nombre d'égalités pourrait occasionner des problèmes dans la façon de répartir les valeurs dans les trois sous-ensembles. Velleman et Hoaglin (1981) proposent, dans leur programme informatique, des solutions à ces problèmes, comme la réduction du nombre de sous-ensembles de trois à deux. Quant à la quantité d'observations qui devrait se retrouver dans les ensembles G et D , lorsqu'il n'y a pas d'égalité parmi les x_i , Hoaglin et al. (1983) proposent trois façons de répartir les observations telles que présentées dans le tableau 1.2.

Plusieurs méthodes peuvent être liées à la méthode de la droite résistante. Ces méthodes ont été développées durant les années 1940 et 1950 dans le but commun de représenter un ensemble de données (x_i, y_i) par une droite (voir Wald 1940, Nair et Shrivastava 1942, Bartlett 1949, Mood 1950 et Kildea 1981). Elles pourraient être considérées comme des

Sous-ensemble	$n = 3m$	$n = 3m + 1$	$n = 3m + 2$
G	$n_G = m$	$n_G = m$	$n_G = m + 1$
C	$n_C = m$	$n_C = m + 1$	$n_C = m$
D	$n_D = m$	$n_D = m$	$n_D = m + 1$

Tableau 1.2 Répartition des observations selon la taille d'échantillon n lorsque toutes les coordonnées x_i des points P_i sont distinctes

précurseurs de la méthode de la droite résistante en ce sens que l'idée est sensiblement la même : on divise les n données en sous-ensembles après avoir ordonné les x_i .

1.4.1 Algorithme

Afin de résoudre l'équation (1.18), on procède par itération. Cette section décrit l'algorithme donné par Velleman et Hoaglin (1981) afin de trouver les estimateurs de α et β . Puisque les auteurs ne proposent pas de méthodes permettant d'estimer le paramètre d'échelle σ , comme ce fut le cas pour les estimateurs-M, nous omettrons dans cette section ce paramètre. L'estimé de β , pour $k \geq 2$, est donné par

$$b_k = b_k^- + (b_k^+ - b_k^-) \left(\frac{e(b_k)^-}{e(b_k)^- - e(b_k)^+} \right), \quad (1.19)$$

où

1. $e_i^{(k)} = y_i - a_k - b_k x_i$
2. $e(b_k) = \text{med}_{(x_i; y_i) \in D}(e_i^{(k)}) - \text{med}_{(x_i; y_i) \in G}(e_i^{(k)})$
3. a) Si $|e(b_{k-1})| < |e(b_{k-2})|$, alors
 - i $e(b_k)^- = e(b_{k-1})$,
 - ii $e(b_k)^+ = e(b_{k-2})$,
 - iii $b_k^- = b_{k-1}$,
 - iv $b_k^+ = b_{k-2}$.

b) Si $|e(b_{k-1})| > |e(b_{k-2})|$, alors

i $e(b_k)^- = e(b_{k-2})$,

ii $e(b_k)^+ = e(b_{k-1})$,

iii $b_k^- = b_{k-2}$,

iv $b_k^+ = b_{k-1}$.

Puisqu'on a $e(b_k)^- < 0 < e(b_k)^+$ ou $e(b_k)^+ < 0 < e(b_k)^-$ et $|e(b_k)^-| < |e(b_k)^+|$, alors on peut voir que $0 < \frac{e(b_k^-)}{e(b_k^-) - e(b_k^+)} < 0,5$. L'algorithme est construit de telle sorte que la valeur a_k n'a pas d'incidence sur l'estimé du paramètre β . Sans perte de généralité, nous posons à cette étape $a_k = 0$, pour tout $k \geq 0$, ce qui donne $e_i^{(k)} = y_i - b_k x_i$.

Les quantités $\text{med}_{(x_i; y_i) \in G}(e_i^{(k)})$ et $\text{med}_{(x_i; y_i) \in D}(e_i^{(k)})$ sont appelées valeurs sommaires. Par exemple, à chaque itération, b_k^- prendra la valeur d'un des estimés b obtenus aux deux itérations précédentes, soit celui dont la différence des valeurs sommaires ($e(b_{k-1})$ ou $e(b_{k-2})$) est la plus près de zéro. Ainsi, pour $k = 2$, si $|e(b_0)| < |e(b_1)|$, alors l'estimé b_2 sera donné par

$$b_2 = b_0 + (b_1 - b_0) \left(\frac{e(b_0)}{e(b_0) - e(b_1)} \right),$$

et si $|e(b_0)| > |e(b_1)|$, alors

$$b_2 = b_1 + (b_0 - b_1) \left(\frac{e(b_1)}{e(b_1) - e(b_0)} \right).$$

L'estimateur de α , lorsqu'il n'est pas fixé à zéro, est obtenu de l'équation

$$a = \text{med}_i(y_i - b x_i),$$

à partir de l'estimé b trouvé par la méthode itérative après convergence. Lorsque $e(b_k)^- = 0$, alors l'estimé b_k de l'équation (1.19) est égal à b_k^- et la convergence est atteinte. En pratique, on appliquera le processus itératif jusqu'à ce que $e(b_k)^-$ soit jugé suffisamment près de zéro. Lorsque $k = 2$, l'utilisation de l'équation (1.19) nécessite la connaissance des valeurs initiales b_0 , b_1 , $e(b_0)$ et $e(b_1)$ afin de déterminer b_2^- , b_2^+ , $e(b_2)^-$ et $e(b_2)^+$. On calcule d'abord b_0 de la façon suivante

$$b_0 = \frac{y_D - y_G}{x_D - x_G},$$

où

$$y_G = \text{med}_{(x_i, y_i) \in G}(y_i), \quad y_D = \text{med}_{(x_i, y_i) \in D}(y_i),$$

et

$$x_G = \text{med}_{(x_i, y_i) \in G}(x_i), \quad x_D = \text{med}_{(x_i, y_i) \in D}(x_i).$$

La valeur $e(b_0)$ est ensuite obtenue par l'équation suivante

$$e(b_0) = \text{med}_{(x_i, y_i) \in D}(e_i^{(0)}) - \text{med}_{(x_i, y_i) \in G}(e_i^{(0)}).$$

Il s'agit de la différence des valeurs sommaires à l'itération $k = 0$. Puis, on calcule l'estimé b_1 comme suit

$$b_1 = b_0 + e(b_0).$$

Enfin, on obtient $e(b_1)$ de la même façon que $e(b_0)$, mais en remplaçant b_0 par b_1 . Les valeurs b_0 et b_1 définissent alors un intervalle contenant la vraie valeur de β . À chaque itération, on restreint l'intervalle en trouvant de nouvelles bornes de façon à se rapprocher de β . Puisque l'algorithme consiste à chercher b_k tel que $e(b_k)^- = 0$ (on voit par l'équation (1.19) qu'à cet instant, il n'y a plus d'amélioration possible à apporter à l'estimé b) on aura généralement des valeurs de $e(b_0)$ et $e(b_1)$ de signes opposés, ce qui est souhaitable puisque cela indique que la vraie valeur β se trouve entre b_0 et b_1 . Si $e(b_0)$ et $e(b_1)$ sont de même signe, alors on ajuste b_1 jusqu'à ce que $e(b_0)$ et $e(b_1)$ soient de signe opposé. Par exemple, si $e(b_0)$ et $e(b_1)$ sont positifs (négatifs), alors on ajoutera un certain nombre positif (négatif) à b_1 de telle sorte que $e(b_1)$ devienne négatif (positif). Cet ajustement sera fait à chaque itération si nécessaire, de façon à ce que $e(b_{k-1})$ et $e(b_k)$ soient toujours de signe opposé, et par conséquent $e(b_{k+1})^-$ et $e(b_{k+1})^+$ aussi.

Afin d'estimer β , il est aussi possible d'itérer avec l'équation $b_k = b_{k-1} + e(b_{k-1})$. Par contre, l'estimé pourrait tendre vers plus d'une valeur (oscillation entre deux valeurs par exemple, voir Hoaglin et al. 1983) ou encore converger très lentement. L'approche donnée par l'équation (1.19) est donc préférable.

Nous allons étudier le comportement de l'estimateur de la droite résistante de Tukey

Groupe	Élément du sous-ensemble
G	$\{P_1, P_2, P_3\} = \{(1,1; 2,4), (1,2; 0,5), (1,5; 7,6)\}$
C	$\{P_4, P_5, P_6, P_7\} = \{(3,0; 10,3), (3,7; 6,4), (3,9; 9,2), (3,9; 10,7)\}$
D	$\{P_8, P_9, P_{10}\} = \{(5,0; 16,6), (5,7; 13,6), (7,1; 18,8)\}$

Tableau 1.3 Répartition des points selon le tableau 1.2 lorsque l'ordonnée du point P_{10} tend vers plus ou moins l'infini

à l'aide des données du tableau 1.1. Nous considérons d'abord le cas où le point P_{10} est clairement aberrant en ordonnée. Les figures 1.37 (a) et 1.37 (b) décrivent le comportement de l'estimateur de la pente en présence du point aberrant $P_{10}^{(I)} = (7,1; 68,8)$ et $P_{10}^{(II)} = (7,1; -32,2)$, respectivement. La pente de la droite ne semble subir que très peu l'influence du point extrême, puisque qu'elle passe par la trajectoire linéaire que décrivent les neuf autres points. Les estimés de β sont donnés, respectivement, par $b_I = 3,3846$ et $b_{II} = 2,4348$, soit une surestimation et une sous-estimation de l'estimé de β lorsque le point P_{10} est retiré de l'ensemble, c'est-à-dire $b_{DR} = 2,9643$. La figure 1.38 indique clairement que l'influence que détient le point aberrant sur la droite est limitée. En effet, on constate que le point P_{10} influence de façon significative l'estimateur de β seulement lorsque son ordonnée se situe dans l'intervalle ouvert $(17; 22,75)$. À l'extérieur de cet intervalle, les estimés de β sont égaux à 3,3846 et 2,4348 lorsque y_{10} tend vers plus ou moins l'infini, respectivement.

Nous allons maintenant tenter d'expliquer le comportement de l'estimateur de β en présence d'une valeur aberrante. La taille de notre échantillon est $n = 10$. Nous avons réparti les données en trois sous-ensembles selon le tableau 1.2, en assignant les points P_6 et P_7 , dont les valeurs x_6 et x_7 sont égales, dans le sous-ensemble C de l'équation (1.16). La répartition des points est présentée dans le tableau 1.3.

À partir de l'équation (1.18), on cherche l'égalité entre les médianes des erreurs des sous-ensembles G et D . Puisque ces deux sous-ensembles possèdent un nombre impair d'observations, la médiane des résidus sera exactement égale à un de ces résidus. Sup-

Domaine de y_{10}	Valeur sommaire de G $\text{med}_{x_i \in G_k}(e_i^{(k)})$	Valeur sommaire de D $\text{med}_{x_i \in D_k}(e_i^{(k)})$	Comportement de b
$(-\infty; 17]$	$e_1^{(k)}$	$e_9^{(k)}$	$b = 2,4348$
$(17; 22,75)$	$e_1^{(k)}$	$e_{10}^{(k)}$	b croît avec y_{10}
$[22,75; \infty)$	$e_1^{(k)}$	$e_8^{(k)}$	$b = 3,3846$

Tableau 1.4 Comportement de l'estimé de β pour les trois points sommaires différents, lorsque l'ordonnée du point P_{10} tend vers plus ou moins l'infini

posons que le processus itératif converge avec une précision suffisante à l'itération k . Puisque nous faisons varier l'ordonnée du point P_{10} tout en gardant fixe son abscisse, on a toujours $x_1 \leq x_2 \leq \dots \leq x_{10}$. La répartition des données selon le tableau 1.3 fait en sorte que les sous-ensembles G et D seront toujours composés des mêmes éléments. Par contre, la valeur sommaire du sous-ensemble D variera dépendamment de la valeur de y_{10} . En fait, pour ce sous-ensemble, on observe trois valeurs sommaires différentes. Ces trois cas sont résumés dans le tableau 1.4.

Dans le premier cas, c'est-à-dire lorsque l'ordonnée du point P_{10} se situe entre moins l'infini et 17, l'estimé de β est la solution de $e_1^{(k)} = e_9^{(k)}$, soit la pente d'une droite passant par les points P_1 et P_9 . Le seul élément variable de l'ensemble D est le point P_{10} . Plus l'ordonnée de ce point s'éloigne vers moins l'infini, plus le résidu $e_{10}^{(k)}$ s'éloigne de $e_9^{(k)}$ et $e_8^{(k)}$. Ces deux derniers étant fixes, on a toujours $e_{10}^{(k)} < e_9^{(k)} \leq e_8^{(k)}$, d'où l'estimation constante de β égal à $b_{II} = 2,4248$. Le deuxième cas correspond à la situation où l'ordonnée du point P_{10} est dans l'intervalle ouvert $(17; 22,75)$. L'estimé de β est la solution de $e_1^{(k)} = e_{10}^{(k)}$, soit la pente d'une droite passant par les points P_1 et P_{10} . La valeur sommaire du sous-ensemble D étant déterminée par $e_{10}^{(k)}$, sa valeur change à mesure que l'ordonnée y_{10} change. L'estimé de β n'est plus constant comme on peut le remarquer sur la figure 1.38. Enfin, la dernière valeur sommaire est obtenue lorsque l'ordonnée du point P_{10} est supérieure à 22,75. L'estimé de β est la solution de $e_1^{(k)} = e_8^{(k)}$, soit la pente d'une droite passant par les points P_1 et P_8 . Le raisonnement est semblable

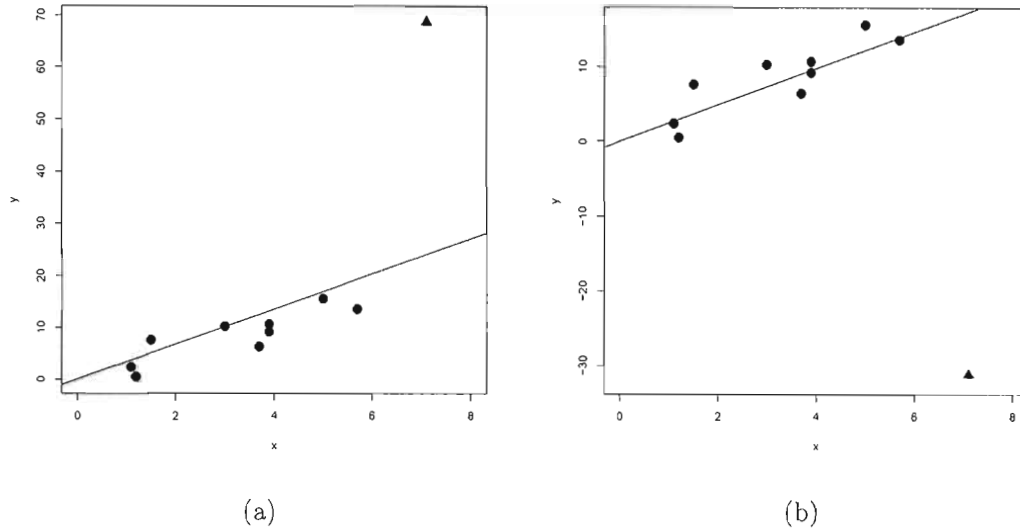


Figure 1.37 Estimé de la droite résistante de Tukey de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7, 1; 68, 8)$, (b) lorsque le point $P_{10}^{(II)} = (7, 1; -31, 2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée.

au premier cas, alors que cette fois-ci, on a toujours $e_9^{(k)} < e_8^{(k)} < e_{10}^{(k)}$. On conclut que la méthode d'estimation produit un estimateur de β robuste à une valeur aberrante en ordonnée. Cependant, l'estimé de β calculé lorsque le point $P_{10} = (7, 1; 22, 75)$ est le même que celui calculé lorsque le point $P_{10} = (7, 1; 100\,000)$. Puisque le point dont les coordonnées sont $(7, 1; 22, 75)$ n'est pas un point aberrant, on s'attend à ce qu'une méthode d'estimation robuste lui accorde plus de poids dans le calcul de l'estimé b , ce qui n'est pas le cas ici. La méthode est donc robuste, mais manque de finesse.

Considérons maintenant le cas où le point P_{10} est aberrant en abscisse. Les figures 1.39 (a) et 1.39 (b) illustrent l'influence que détiennent les points $P_{10}^{(III)} = (57, 1; 18, 8)$ et $P_{10}^{(IV)} = (-42, 9; 18, 8)$, respectivement. Les estimés de β sont donnés par $b_{III} = 2,4348$ et $b_{IV} = 2,9643$, respectivement. La méthode d'estimation de la droite résistante nécessite qu'on ordonne les valeurs x_i , de la plus petite à la plus grande. Puisqu'on fait varier l'abscisse du point P_{10} , l'ordre parmi les x_i en sera affecté. La répartition des

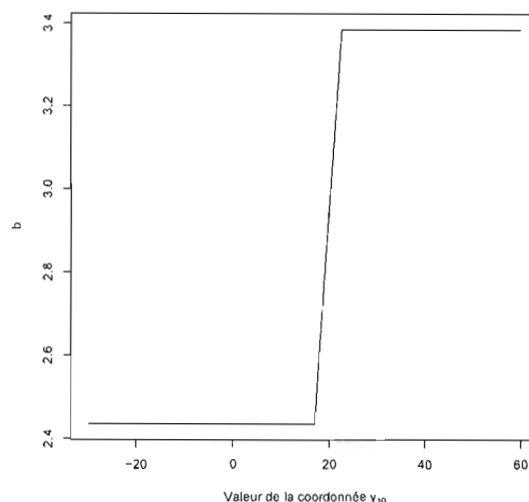


Figure 1.38 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} .

observations dans les sous-ensembles G, C et D se fait de différentes façons à mesure que la valeur aberrante parcourt les réels. Afin de maintenir l'égalité des tailles des groupes G et D ($n_G = n_D$), tout en s'assurant que les valeurs x_i égales entre elles se retrouvent dans le même groupe, nous avons réparti les observations entre les sous-ensembles G, C et D , selon les différents cas illustrés dans le tableau 1.5. La figure 1.40 décrit le comportement de l'estimé b lorsque le point P_{10} tend vers plus ou moins l'infini en abscisse. Nous verrons que les différents sauts qu'on peut y observer sont dus à la répartition des observations décrite dans le tableau 1.5.

Supposons que la convergence de l'estimé b atteint une précision suffisante à l'itération k . On considère à nouveau les sous-ensembles G et D des équations (1.15) et (1.17). Pour le cas 1, c'est-à-dire, lorsque $x_{10} \in (-\infty; 3)$, les sous-ensembles G et D contiennent chacun quatre éléments. Tous les points des deux sous-ensembles ont des coordonnées fixes à l'exception du point P_{10} dont le résidu $e_{10}^{(k)}$ devient de plus en plus grand, à mesure que x_{10} s'éloigne vers moins l'infini. Le résidu $e_{10}^{(k)}$ sera toujours plus grand que

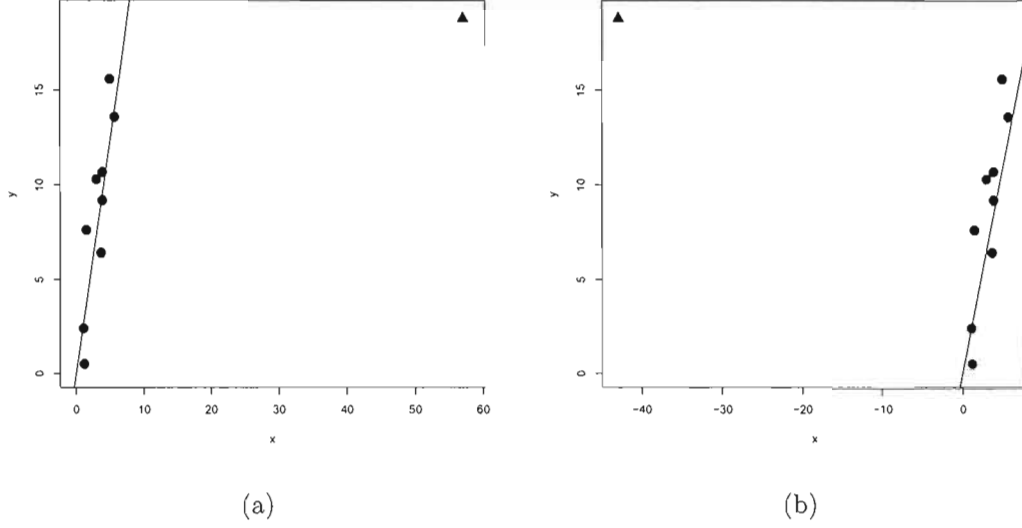


Figure 1.39 Estimé de la droite résistante de Tukey de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57, 1; 18, 8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42, 9; 18, 8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée x a été modifiée.

les autres résidus de telle sorte que nous avons $e_2^{(k)} < e_1^{(k)} < e_3^{(k)} < e_{10}^{(k)}$. Ainsi, la valeur sommaire de G est toujours déterminée par $\frac{1}{2}(e_1^{(k)} + e_3^{(k)})$. Dans le sous-ensemble D , les résidus sont ordonnés tels que $e_6^{(k)} < e_9^{(k)} < e_7^{(k)} < e_8^{(k)}$. La valeur sommaire de D est donnée par $\frac{1}{2}(e_7^{(k)} + e_9^{(k)})$. On voit clairement que la valeur aberrante n'a aucune influence sur le point sommaire dans cet intervalle. L'effet de P_{10} sur l'estimation de β sera donc le même pour toute valeur de x_{10} dans l'intervalle. L'estimé $b_k = 2,0429$ pour tout x_{10} dans l'intervalle est la solution de

$$\begin{aligned}
 e(b_k) &= \text{med}_{(x_i, y_i) \in G}(e_i^{(k)}) - \text{med}_{(x_i, y_i) \in D}(e_i^{(k)}) \\
 &= \frac{1}{2}(e_1^{(k)} + e_3^{(k)}) - \frac{1}{2}(e_7^{(k)} + e_9^{(k)}) \\
 &= 0.
 \end{aligned}$$

Il s'agit donc de la pente d'une droite passant entre P_1 et P_3 et entre P_7 et P_9 de façon à ce que la distance verticale entre la droite et les quatre points soit la même.

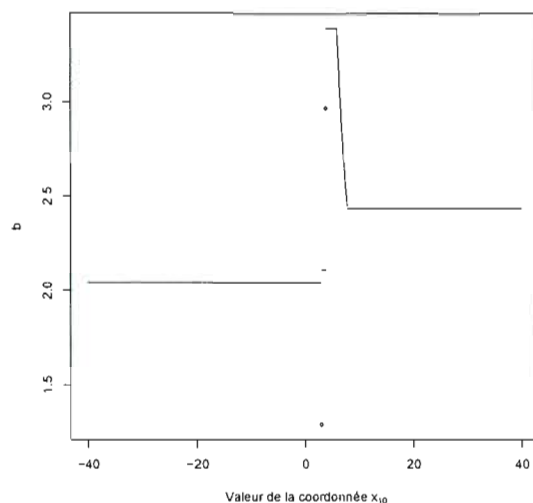


Figure 1.40 Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} .

Cas	Domaine x_{10}	Groupe G	Groupe C	Groupe D
1	$(-\infty; 3)$	P_{10}, P_1, P_2, P_3	P_4, P_5	P_6, P_7, P_8, P_9
2	3	$P_{10}, P_1, P_2, P_3, P_4$	\emptyset	P_5, P_6, P_7, P_8, P_9
3	$(3; 3,9)$	P_1, P_2, P_3, P_4	P_5, P_{10}	P_6, P_7, P_8, P_9
4	3,9	P_1, P_2, P_3, P_4, P_5	\emptyset	$P_{10}, P_6, P_7, P_8, P_9$
5	$(3,9; \infty)$	P_1, P_2, P_3	P_4, P_5, P_6, P_7	P_8, P_9, P_{10}

Tableau 1.5 Répartition des données originales selon la position de l'abscisse du point P_{10} par rapport aux autres coordonnées x_i , pour $i = 1, \dots, 9$

Dans le deuxième cas, $x_{10} = x_4 = 3$. De plus, on a toujours $x_6 = x_7 = 3,9$. Nous avons réparti les données de façon à ce que les sous-ensembles G et D aient chacun cinq éléments. Nous avons séparé le jeu de données en deux groupes plutôt que trois dans le but d'avoir des tailles égales pour les sous-ensembles G et D et aussi pour s'assurer que les valeurs x égales se retrouvent dans le même sous-ensemble. Pour le sous-ensemble G , on observe que $e_2^{(k)} < e_1^{(k)} < e_3^{(k)} < e_4^{(k)} < e_{10}^{(k)}$ alors que pour le sous-ensemble D , on a $e_5^{(k)} < e_6^{(k)} < e_7^{(k)} < e_9^{(k)} < e_8^{(k)}$. Ainsi, les valeurs sommaires des deux sous-ensembles sont données par $e_3^{(k)}$ et $e_7^{(k)}$, respectivement, d'où l'estimé de β égal à 1,2917, soit la pente d'une droite passant par les points P_3 et P_7 .

Dans le troisième cas, c'est-à-dire lorsque $x_{10} \in (3; 3,9)$, les groupes G et D contiennent à nouveau quatre éléments chacun. Par contre, le seul élément variable, c'est-à-dire le point P_{10} , ne se retrouve ni dans G , ni dans D . Tous les autres éléments étant fixes, les points sommaires des deux sous-ensembles G et D le seront aussi pour toute valeur de x_{10} dans l'intervalle. Les valeurs sommaires $\frac{1}{2} (e_1^{(k)} + e_4^{(k)})$ et $\frac{1}{2} (e_7^{(k)} + e_9^{(k)})$ des sous-ensembles G et D respectivement, sont à l'origine de l'estimation constante de β égale à 2,1091.

Lorsque $x_{10} = x_6 = x_7 = 3,9$, les dix observations sont à nouveau réparties en deux groupes de cinq éléments en respectant la condition que x_6, x_7 et x_{10} soient dans le même sous-ensemble. La valeur sommaire de G est donnée par $e_1^{(k)}$ alors que celle du sous-ensemble D est $e_7^{(k)}$. L'estimé $b = 2,9643$ est donc la valeur de β telle que $e_1^{(k)} = e_7^{(k)}$, soit la pente d'une droite passant par les points P_1 et P_7 .

Pour le dernier intervalle, $(3,9; \infty)$, l'estimé b se comporte différemment selon que x_{10} soit dans l'intervalle $I_1 = (3,9; 5,95)$, $I_2 = [5,95; 7,85)$ et $I_3 = [7,85; \infty)$. Les sous-ensembles sont donnés par $G = \{P_1, P_2, P_3\}$ et $D = \{P_8, P_9, P_{10}\}$. Les différentes valeurs de b obtenues (b constant égale à 3,3846 lorsque $x_{10} \in I_1$, b décroissant lorsque $x_{10} \in I_2$ et b constant égale à 2,4348 lorsque $x_{10} \in I_3$), sont dues aux différentes valeurs sommaires résultant de l'algorithme. Pour des valeurs de x_{10} dans l'intervalle I_1 , $\text{med}_{(x_i, y_i) \in G}(e_{i_k}) = e_1^{(k)}$ alors que $\text{med}_{(x_i, y_i) \in D}(e_{i_k}) = e_8^{(k)}$. L'estimé $b = 3,3846$ est la valeur de β qui est solution de $e_1^{(k)} = e_8^{(k)}$, soit la pente d'une droite passant par les points P_1 et P_8 . Pour

Cas	Domaine de x_{10}	Valeur sommaire de G $\text{med}_{(x_i, y_i) \in G}(e_{i_k})$	Valeur sommaire de D $\text{med}_{(x_i, y_i) \in D}(e_{i_k})$	b
1	$(-\infty; 3)$	$\frac{1}{2}(e_1^{(k)} + e_3^{(k)})$	$\frac{1}{2}(e_7^{(k)} + e_9^{(k)})$	2,0429
2	3	$e_3^{(k)}$	$e_7^{(k)}$	1,2917
3	$(3; 3,9)$	$\frac{1}{2}(e_1^{(k)} + e_4^{(k)})$	$\frac{1}{2}(e_7^{(k)} + e_8^{(k)})$	2,1091
4	3,9	$e_1^{(k)}$	$e_7^{(k)}$	2,9643
5	$(3,9; 5,95)$	$e_1^{(k)}$	$e_8^{(k)}$	3,3846
6	$[5,95; 7,85)$	$e_1^{(k)}$	$e_{10}^{(k)}$	décroissant
7	$[7,85; \infty)$	$e_1^{(k)}$	$e_9^{(k)}$	2,4348

Tableau 1.6 Valeurs sommaires des groupes G et D selon la valeur de la coordonnée x_{10} . Les valeurs sommaires de G et D sont définies par $\text{med}_{(x_i, y_i) \in G}(e_i^{(k)})$ et $\text{med}_{(x_i, y_i) \in D}(e_i^{(k)})$, respectivement

des valeurs de x_{10} dans l'intervalle I_2 , les valeurs sommaires des sous-ensembles G et D sont données par les résidus $e_1^{(k)}$ et $e_{10}^{(k)}$, respectivement. Comme la pente est celle d'une droite passant par les points P_1 et P_{10} , et que le point P_{10} est variable, cela explique que la pente diminue à mesure que x_{10} augmente. Enfin, pour des valeurs de x_{10} dans l'intervalle I_3 , l'estimé $b = 2,4348$ est la valeur de β telle que les valeurs sommaires $e_1^{(k)}$ de G et $e_9^{(k)}$ de D sont égales, soit la pente d'une droite passant par les points P_1 et P_9 .

Il en résulte que l'estimateur du paramètre β par la méthode de la droite résistante est robuste à une valeur aberrante en abscisse. On constate comme dans le cas d'une valeur aberrante en ordonnée que cette méthode d'estimation manque de finesse. Par exemple, la valeur de l'estimé de β est la même pour tout point P_{10} dont $x_{10} \in [7,85; \infty)$. Compte tenu du fait que le point $(7,9; 18,8)$, dont les coordonnées s'apparentent à celles du point P_{10} original, n'est pas une donnée aberrante, on s'attend à ce que son influence sur le calcul de l'estimé de b soit plus important que pour un point dont x_{10} tend vers l'infini. La figure 1.40 nous indique plutôt le contraire. Le tableau 1.6 résume le comportement de l'estimé b , présenté sur la figure 1.40, en fonction des valeurs sommaires.

En terminant, rappelons que l'application de la méthode d'estimation de Tukey à un modèle de la forme

$$y_i = \beta x_i + \varepsilon_i$$

doit être faite avec prudence. Poser $\alpha = 0$ ne causera pas de problème si l'estimé de α est près de zéro. Dans le cas contraire, la pente de la droite pourrait être mal estimée. Par exemple, supposons que nous ajoutons la quantité 1.5 à chaque coordonnée x_i des points du tableau 1.1, de telle sorte que les coordonnées des points P_1, \dots, P_{10} soient maintenant $(2.6; 2.4), \dots, (8.6; 18.8)$. Alors, la figure 1.41 illustre l'estimation par la méthode de la droite résistante des modèles $y_i = \beta x_i + \varepsilon_i$ (la droite noire) et $y_i = \alpha + \beta x_i + \varepsilon_i$ (la droite rouge). Pour le second modèle, l'estimé de la pente est donné par 2,7333 alors que l'estimé de α est égal à -4,7067. Pour le modèle dont l'ordonnée à l'origine est nulle, puisque α n'intervient pas dans l'estimation de b , alors le paramètre β est aussi estimé à 2,7333. On remarque que la droite noire passe au dessus des points de l'ensemble, ce qui n'est naturellement pas souhaitable. Il devient clair que l'estimé 2,7333 du paramètre β du modèle $y_i = \beta x_i + \varepsilon_i$ est surévalué dû au fait qu'on fixe $\alpha = 0$ dans une méthode qui estime α et β à la fois.

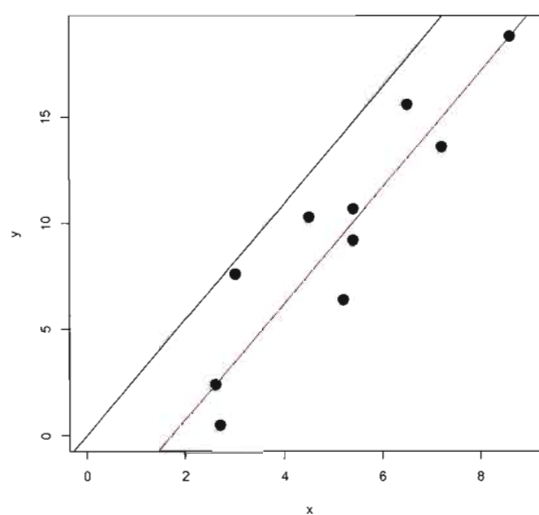


Figure 1.41 Droites estimées par la méthode de la droite résistante pour le modèle $y_i = \beta x_i + \varepsilon_i$ (en noir) et pour le modèle $y_i = \alpha + \beta x_i + \varepsilon_i$ (en rouge).

CHAPITRE II

MÉTHODE D'ESTIMATION BAYÉSIENNE AVEC PARAMÈTRE D'ÉCHELLE CONNU

2.1 Contexte théorique

Dans ce chapitre, nous proposons une méthode d'estimation robuste dans un contexte bayésien et basée sur des distributions à ailes relevées (*heavy tailed*). Au chapitre précédent, nous avons énuméré quelques hypothèses sur les erreurs ε_i , entre autres que $\frac{\varepsilon_i}{\sigma}$ i.i.d F , pour une distribution quelconque F . Dans la littérature, on suppose souvent cette distribution F comme étant normale (par exemple, voir Lange et al. 1989, Lawrence et Arthur 1990, Birkes et Dodge 1993, Rousseeuw et Leroy 2003). Lorsque cette hypothèse n'est pas satisfaite, l'inférence statistique s'en trouve affectée. Parmi les solutions permettant de minimiser l'impact que peut provoquer le non respect de l'hypothèse de normalité des erreurs, l'une d'elle consiste à formuler une nouvelle hypothèse quant à la distribution des erreurs. Afin de caractériser les erreurs, aux fins de robustesse, on favorisera une densité ayant des ailes plus épaisses que celle de la loi normale.

2.2 Contexte bayésien

On considère un échantillon de variables aléatoires X_1, \dots, X_n . Pour un paramètre θ , on suppose que les variables $X_i|\theta$, pour $i = 1, \dots, n$, sont conditionnellement indépendantes entre elles. La densité conditionnelle de chaque variable $X_i|\theta$ est donnée par $f_i(x_i|\theta)$,

pour $\theta \in \Theta$ et $X_i \in \Omega$, où Θ et Ω sont des sous-ensembles de \mathbb{R} . L'approche bayésienne consiste à combiner l'information contenue dans l'échantillon de données à de l'information complémentaire, appelée *a priori*. Cette information *a priori* provient souvent d'expériences ou d'opinions personnelles et est indépendante des données observées. L'information *a priori* est généralement spécifiée avant de recueillir l'échantillon de données en associant au paramètre θ une loi. La mise à jour de l'information *a priori* se fait en y ajoutant l'information contenue dans les données par le théorème de Bayes, qui stipule que $\Pr[A|B] = \Pr[B|A] \Pr[A] / \Pr[B]$, pour A et B , deux sous-espaces d'un espace échantillonnal S . Ainsi, l'application de ce théorème nous permet d'obtenir la densité *a posteriori* de θ , que nous notons $\pi(\theta|\underline{x})$, par

$$\pi(\theta|\underline{x}) = \frac{\pi(\theta) \prod_{i=1}^n f_i(x_i|\theta)}{m(\underline{x})},$$

où $\pi(\theta)$ est la densité *a priori*, $\underline{x} = (x_1, x_2, \dots, x_n)$ et $m(\underline{x}) = \int_{\Theta} \pi(\theta) \prod_{i=1}^n f_i(x_i|\theta) d\theta$ est la densité marginale de \underline{X} . En statistique bayésienne, le processus d'expérimentation est supposé se faire en deux étapes. Tout d'abord, on considère la variable aléatoire θ dont la distribution est donnée par $\pi(\theta)$. De cette distribution, une valeur de θ est tirée de façon aléatoire. On obtient alors θ^{obs} , une réalisation ou une observation de la variable aléatoire θ . Sachant la valeur observée θ^{obs} , la deuxième étape consiste à tirer au hasard, un échantillon de variables aléatoires X_1, \dots, X_n dont chaque variable provient d'une distribution $f_i(x_i|\theta^{obs})$ notée généralement par $f_i(x_i|\theta)$. Ainsi, chaque nouvelle expérimentation produit un nouveau θ et un nouvel échantillon x_1, \dots, x_n . En statistique classique (fréquentiste), la première étape n'est pas considérée. L'expérimentation classique consiste donc à tirer au hasard, un échantillon de variables aléatoires X_1, \dots, X_n distribuées selon $f_i(x_i|\theta)$, où θ est inconnu mais considéré fixe. Dans ce cas, chaque nouvelle expérimentation produit un nouvel échantillon de variables x_1, \dots, x_n , mais le paramètre θ est toujours le même puisqu'il est fixe. En statistique bayésienne, il est possible de faire de l'inférence à partir de la densité *a posteriori* de θ , $\pi(\theta|\underline{x})$, et de la densité prédictive d'une nouvelle observation X_{n+1} , soit $f(x_{n+1}|\underline{x})$.

2.3 Le problème de position

Au premier chapitre, nous avons exposé différentes méthodes de régression robuste afin d'estimer la pente de la droite de l'équation

$$y_i = \beta x_i + \varepsilon_i. \quad (2.1)$$

Dans cette section, nous allons formuler la densité *a posteriori* de β et du coup, nous allons voir que le problème d'estimation de la pente d'une droite de régression peut se traduire en un problème d'estimation d'un paramètre de position. La motivation d'écrire ce problème sous cette forme provient principalement du fait que les résultats théoriques de robustesse bayésienne que nous verrons à la section 2.5 s'appliquent au problème d'estimation d'un paramètre de position. La contribution principale de ce mémoire réside dans ce lien.

Soit un ensemble de données (x_i, y_i) , pour $i = 1, \dots, n$, pour lequel nous faisons l'hypothèse qu'il existe une relation linéaire, représentée par l'équation (2.1), entre les variables X_i et Y_i . Nous considérons que les variables aléatoires X_i , pour $i = 1, \dots, n$, ont une distribution quelconque qui ne dépend pas de β . Cette distribution n'aura alors aucun impact sur l'inférence *a posteriori* de β . Puisque nous travaillons dans un contexte bayésien, nous considérons aussi les variables Y_i et ε_i , de même que le paramètre β de l'équation (2.1), comme des variables aléatoires. Nous supposons que les variables ε_i sont indépendamment et identiquement distribuées. Généralement, on choisit la densité des erreurs ε_i symétrique, définie sur \mathbb{R} et centrée en zéro. Cette densité, évaluée au point ε_i est notée $f_\varepsilon(\varepsilon_i)$. Nous avons omis le paramètre d'échelle σ , puisque étant considéré connu, il est implicitement inclus dans $f_\varepsilon(\varepsilon_i)$. La densité *a posteriori* du paramètre β est alors donnée par

$$\begin{aligned} \pi(\beta | \underline{x}, \underline{y}) &= \frac{\pi(\beta) f_{X,Y}(\underline{x}, \underline{y} | \beta)}{m(\underline{x}, \underline{y})} \\ &\propto \pi(\beta) f_Y(\underline{y} | \underline{x}, \beta) f_X(\underline{x} | \beta) \\ &\propto \pi(\beta) f_Y(\underline{y} | \underline{x}, \beta) f_X(\underline{x}) \\ &\propto \pi(\beta) f_Y(\underline{y} | \underline{x}, \beta). \end{aligned}$$

Puisque $Y_i = \beta X_i + \varepsilon_i$, et que ε_i sont indépendantes entre elles, on voit bien que $Y_i|X_i, \beta$ sont aussi indépendantes pour $i = 1 \dots, n$. Par conséquent, on a

$$\pi(\beta|\underline{x}, \underline{y}) \propto \pi(\beta) \prod_{i=1}^n f_{Y_i}(y_i|x_i, \beta).$$

Par les techniques de changement de variables, puisque $\varepsilon_i = Y_i - \beta X_i$, chaque variable Y_i a comme densité conditionnelle, sachant X_i et β ,

$$\begin{aligned} f_{Y_i}(y_i|x_i, \beta) &= f_{\varepsilon}(y_i - \beta x_i) \left| \frac{d\varepsilon_i}{dY_i} \right| \\ &= f_{\varepsilon}(y_i - \beta x_i). \end{aligned}$$

On a alors

$$\begin{aligned} \pi(\beta|\underline{x}, \underline{y}) &\propto \pi(\beta) \prod_{i=1}^n f_{Y_i}(y_i|x_i, \beta) \\ &\propto \pi(\beta) \prod_{i=1}^n f_{\varepsilon}(y_i - \beta x_i). \end{aligned}$$

Si on définit la variable aléatoire $Z_i = Y_i/X_i$, ce qui est équivalent à $\varepsilon_i = \frac{Z_i - \beta}{1/X_i}$, nous pourrions alors considérer β comme un paramètre de position pour les variables aléatoires Z_i , considérant X_1, \dots, X_n connus. En effet, la densité de Z_i , sachant X_i et β , est

$$\begin{aligned} g_{Z_i}(z_i|x_i, \beta) &= f_{\varepsilon}\left(\frac{z_i - \beta}{1/x_i}\right) \left| \frac{d\varepsilon_i}{dZ_i} \right| \\ &= \frac{1}{1/x_i} f_{\varepsilon}\left(\frac{z_i - \beta}{1/x_i}\right). \end{aligned} \quad (2.2)$$

On note que les densités de $Y_i|X_i, \beta$ ne diffèrent que par un paramètre de position (βx_i), tandis que les densités de $Z_i|X_i, \beta$ ne diffèrent que par un paramètre d'échelle ($1/x_i$), et partagent le même paramètre de position β . Si on considère les observations $z_1, \dots, z_n = y_1/x_1, \dots, y_n/x_n$, on a alors

$$\begin{aligned} \pi(\beta|\underline{z}, \underline{z}) &\propto \pi(\beta) \prod_{i=1}^n g_{Z_i}(z_i|x_i, \beta) \\ &\propto \pi(\beta) \prod_{i=1}^n \frac{1}{1/x_i} f_{\varepsilon}\left(\frac{z_i - \beta}{1/x_i}\right), \end{aligned}$$

puisque $Z_i|X_i, \beta$ sont conditionnellement indépendantes, du fait que $Y_i|X_i, \beta$ le sont aussi et que $Z_i = Y_i/X_i$, pour $i = 1, \dots, n$.

Si nous voulons faire le lien avec la notation utilisée dans Desgagné et Angers (2007) concernant leur résultat de robustesse avec paramètre de position, on n'a qu'à considérer implicitement X_1, \dots, X_n connues et poser

$$f_i(z_i - \beta) = g_{Z_i}(z_i|x_i, \beta) = \frac{1}{1/x_i} f_\varepsilon\left(\frac{z_i - \beta}{1/x_i}\right),$$

où les densités f_i ne diffèrent que par un paramètre d'échelle $1/x_i$ connu. On est bien en présence d'un problème d'estimation d'un paramètre de position avec $Z_i|\beta$ conditionnellement indépendantes ayant comme densité $f_i(z_i - \beta)$, et la densité *a priori* de β est encore $\pi(\beta)$ qu'on note aussi $f_0(z_0 - \beta)$, où z_0 est un paramètre de position connu. Ainsi, on a

$$\pi(\beta|z) \propto \prod_{i=0}^n f_i(z_i - \beta). \quad (2.3)$$

2.4 Aperçu des avancées en robustesse bayésienne

Lorsque les différentes sources d'information sont conflictuelles (c'est-à-dire lorsqu'il y a présence de valeurs aberrantes dans l'échantillon de données et/ou lorsque la distribution *a priori* est incorrectement spécifiée), la distribution *a posteriori*, calculée à partir d'une procédure d'estimation bayésienne non robuste, aura tendance à faire un compromis non réaliste entre les deux sources d'information. Différentes approches ont donc été proposées afin de juger de la source d'information la plus fiable lorsqu'il y a conflit. Afin de restreindre l'effet de valeurs aberrantes sur la densité *a posteriori*, Lindley (1968) a suggéré l'emploi de distributions à ailes relevées. Suivant cette idée, Dawid (1973) a établi des conditions sous lesquelles une source d'information conflictuelle sera écartée à mesure qu'elle s'éloigne de la source d'information non conflictuelle. Principalement, ces conditions stipulent que la source d'information conflictuelle devrait être associée à une distribution ayant des ailes relevées. Hill (1974) et O'Hagan (1979) ont énoncé des conditions plus faciles à vérifier que celles de Dawid. La notion de crédence, caractérisant les ailes d'une densité, a été introduite par O'Hagan (1990) pour des densités dont le

comportement dans les ailes s'apparente à des distributions comme celle de la Student. Angers (2000) a généralisé le concept de crédence, définissant la p-crédence pour un plus grand nombre de lois. Récemment, Desgagné et Angers (2007) ont énoncé des conditions plus générales, basées entre autres sur la p-crédence, permettant de calculer des densités *a posteriori* robustes. Les conditions de robustesse concernent essentiellement l'épaisseur des ailes d'une densité, qui doivent être suffisamment relevées. Aussi, les sources d'information privilégiées en cas de conflit doivent avoir une densité dont les ailes sont moins relevées que les densités des sources conflictuelles.

2.5 Notions sur les densités à ailes relevées

Dans un contexte bayésien, les densités ayant des ailes relevées s'avèrent être d'une grande utilité pour faire de l'inférence robuste. Dans le but d'y parvenir, on doit donc être en mesure de pouvoir choisir ces densités de façon efficace. Le choix de ces densités doit être soumis à des conditions d'épaisseur des ailes et de régularité que doivent satisfaire leurs ailes si on désire présenter des résultats robustes. Dans cette section, nous énumérerons dans un premier temps les conditions d'épaisseur et de régularité permettant le rejet de l'information conflictuelle. Nous verrons par la suite que ces conditions peuvent être simplifiées grâce à la notion de p-crédence.

2.5.1 Conditions d'épaisseur et de régularité pour les ailes d'une densité

Considérons le contexte bayésien de la section 2.2. Rappelons que nous supposons que les variables $X_i|\theta$, pour $i = 1, \dots, n$ sont conditionnellement indépendantes entre elles, où θ est un paramètre de position. Soit $\underline{x}_n = (x_1, \dots, x_n)$, un vecteur échantillonnal composé de n observations tirées d'une population ayant comme densité $f_i(x_i - \theta)$. Afin de mettre l'emphase sur le fait que la densité *a priori* constitue une autre source d'information, nous noterons la distribution du paramètre θ par $f_0(x_0 - \theta)$ plutôt que par la notation usuelle $\pi(\theta - x_0)$, où x_0 est un paramètre de position connu. Parmi les $n + 1$ sources d'information (x_0, x_1, \dots, x_n) , on en considère $k + 1$ fixées, notées

par $(x_{j_0}, x_{j_1}, \dots, x_{j_k})$ où (j_0, j_1, \dots, j_n) est une permutation quelconque de l'ensemble $(0, 1, \dots, n)$ et où $0 \leq k \leq n$. Les $n - k$ sources restantes sont considérées comme étant des sources d'information conflictuelles (l'échantillon de données x_n contient des valeurs aberrantes, ou le paramètre de position de la distribution *a priori* est incorrectement spécifié). Parmi les sources d'information conflictuelles, on considère que $m - k$ d'entre elles tendent vers moins l'infini, notées par $(x_{j_{k+1}}, \dots, x_{j_m})$ et que les $n - m$ restantes tendent vers plus l'infini, notées $(x_{j_{m+1}}, \dots, x_{j_n})$ où $0 \leq k \leq m \leq n$. Finalement, soit $\phi = (-x_{j_{k+1}}, \dots, -x_{j_m}, x_{j_{m+1}}, \dots, x_{j_n})$, le vecteur contenant l'information conflictuelle. Nous supposons que la densité f est une densité propre (c'est-à-dire $f(x) \geq 0$, $\forall x \in \mathbb{R}$ et $\int_{-\infty}^{\infty} f(x)dx = 1$), positive et bornée. Alors les conditions que doivent respecter les densités $f_i(x_i - \theta)$, pour $i = 0, \dots, n$, afin de pouvoir faire de l'inférence robuste, telles que données dans Desgagné et Angers (2007), sont :

C1 : $\forall \varepsilon > 0, \forall h > 0$, il existe une constante $A_1(\varepsilon, h) > 0$ telle que $z > A_1(\varepsilon, h)$ ($z < -A_1(\varepsilon, h)$ pour l'aile à gauche) et $|\theta| < h \Rightarrow 1 - \varepsilon \leq \frac{f(z+\theta)}{f(z)} \leq 1 + \varepsilon$.

Pour les conditions C2 et C3, il existe des constantes $A_2 > 0$ et $M_2 > 1$ et des densités propres f^* et g telles que pour tout $z > A_2$ ($z < -A_2$ pour l'aile à gauche),

C2 : $\frac{f^2(z/2)}{f(z)g(z/2)} \leq M_2$,

C3 : $\frac{d^2}{dz^2} \log f^*(z) \geq \frac{d^2}{dz^2} \log g(z) \geq 0$,

où f^* peut être choisie de telle sorte que le comportement de ses ailes est semblable à celui de la densité f . De façon plus formelle, la densité f^* doit satisfaire la condition suivante : il existe des constantes $B > 0$ et $0 < K_1 < K_2 < \infty$ de tel sorte que $z > B$ ($z < -B$ pour l'aile à gauche) $\Rightarrow K_1 \leq \frac{f(z)}{f^*(z)} \leq K_2$.

Les conditions C1 et C2 stipulent que les ailes de la densité sont suffisamment relevées. En particulier, la condition C1 affirme que pour une valeur de z assez grande, un changement de position apporté sur la variable aléatoire Z n'a pas d'impact sur le comportement de l'aile de la densité. La condition C3 affirme que pour une valeur assez grande de z , les ailes de la densité sont convexes (condition de régularité). Sous

ces conditions, le théorème suivant, de Desgagné et Angers (2007), nous permet de déterminer le comportement robuste de la densité *a posteriori*.

Théorème 1. *Pour tous nombres entiers k et m tel que $0 \leq k \leq m \leq n$ et pour toutes valeurs constantes $x_{j_0}, x_{j_1}, \dots, x_{j_k}$, si les conditions C1 à C3 sont satisfaites pour les ailes à gauche des densités $f_{j_{k+1}}, \dots, f_{j_m}$ et pour les ailes à droite des densités $f_{j_{m+1}}, \dots, f_{j_n}$, et si*

$$\lim_{\theta \rightarrow -\infty} \frac{\prod_{i=0}^k f_{j_i}(x_{j_i} - \theta)}{\prod_{i=k+1}^m f_{j_i}(\theta)} = 0 \text{ lorsque } k < m, \text{ et} \quad (2.4)$$

$$\lim_{\theta \rightarrow \infty} \frac{\prod_{i=0}^k f_{j_i}(x_{j_i} - \theta)}{\prod_{i=m+1}^n f_{j_i}(\theta)} = 0 \text{ lorsque } m < n, \quad (2.5)$$

alors

a) $\theta|_{\underline{x}_n} \xrightarrow{\mathcal{L}} \theta|_{\underline{x}_k}$ lorsque $\phi \rightarrow \infty$, où les densités des variables aléatoires $\theta|_{\underline{x}_n}$ et $\theta|_{\underline{x}_k}$ évaluées en θ sont données par $\pi(\theta|_{\underline{x}_n})$ et $\pi(\theta|_{\underline{x}_k})$.

De plus, pour toute fonction $w(\cdot)$ définie sur \mathbb{R} telle que $\mathbb{E}^{\pi(\theta|_{\underline{x}_k})}[|w(\theta)|] < \infty$ et $|w(\theta)|\pi(\theta|_{\underline{x}_k})$ est bornée supérieurement, si

$$\lim_{\theta \rightarrow -\infty} \frac{w(\theta) \prod_{i=0}^k f_{j_i}(x_{j_i} - \theta)}{\prod_{i=k+1}^m f_{j_i}(\theta)} = 0 \text{ lorsque } k < m, \text{ et} \quad (2.6)$$

$$\lim_{\theta \rightarrow \infty} \frac{w(\theta) \prod_{i=0}^k f_{j_i}(x_{j_i} - \theta)}{\prod_{i=m+1}^n f_{j_i}(\theta)} = 0 \text{ lorsque } m < n, \quad (2.7)$$

alors

b) $\lim_{\phi \rightarrow \infty} \mathbb{E}^{\pi(\theta|_{\underline{x}_n})}[w(\theta)] = \mathbb{E}^{\pi(\theta|_{\underline{x}_k})}[w(\theta)]$.

La condition de l'équation (2.4) stipule que l'aile à gauche de la densité proportionnelle au produit de densités $\prod_{i=k+1}^m f_{j_i}(\theta)$ est plus relevée que celle de la densité de θ proportionnelle au produit de densités $\prod_{i=0}^k f_{j_i}(x_{j_i} - \theta)$. De façon équivalente, la condition de l'équation (2.5) stipule que l'aile à droite de la densité proportionnelle au produit de densités $\prod_{i=m+1}^n f_{j_i}(\theta)$ est plus relevée que celle de la densité de θ proportionnelle au produit de densités $\prod_{i=0}^k f_{j_i}(x_{j_i} - \theta)$.

Sous les conditions C1 à C3 ainsi que sous les conditions des équations (2.4) et (2.5), le résultat a) indique que lorsque les sources d'information conflictuelles divergent à une

vitesse quelconque, la distribution *a posteriori* (sachant les n observations) converge vers la distribution *a posteriori* dont on a retiré ces sources d'information conflictuelles. Concrètement, le théorème affirme que la contribution des différentes sources d'information conflictuelles dans le calcul de la distribution *a posteriori* diminue à mesure que ϕ s'éloigne vers l'infini. Si on ajoute les conditions des équations (2.6) et (2.7), alors par le résultat b), le théorème indique entre autres que les moments *a posteriori* de la distribution $\pi(\theta|x_1, \dots, x_n)$ convergent vers les moments *a posteriori* de la distribution $\pi(\theta|x_1, \dots, x_k)$, c'est-à-dire la distribution *a posteriori* dont on a retiré les sources d'information conflictuelles.

2.5.2 Famille de densités GEP et p-crédence

Il est possible de remplacer les conditions énoncées dans le théorème 1 par des conditions plus faciles à vérifier, basées sur la p-crédence. Cependant, cette simplification entraîne une restriction quant aux densités f pouvant être utilisées. En effet, bien que la p-crédence soit définie pour un très grand nombre de lois connues, il en existe quelques unes pour lesquelles elle n'existe pas. Desgagné et Angers (2005) caractérisent, à l'aide de la p-crédence, le comportement des ailes d'une densité en la comparant à la densité GEP (*generalized exponential power*). La forme générale de la densité GEP telle que spécifiée par Desgagné et Angers (2005) est donnée par

$$\begin{aligned}
 p(z|\gamma, \delta, \alpha, \lambda, z_0) &= K(\gamma, \delta, \alpha, \lambda, z_0) \exp \{ -\delta \max(|z|, z_0)^\gamma \} \\
 &\quad \times \max(|z|, z_0)^{-\alpha} \log^{-\lambda} [\max(|z|, z_0)] \\
 &\propto \begin{cases} e^{-\delta|z|^\gamma} |z|^{-\alpha} \log^{-\lambda} |z|; & \text{si } |z| > z_0, \\ e^{-\delta z_0^\gamma} z_0^{-\alpha} \log^{-\lambda} z_0; & \text{si } |z| \leq z_0, \end{cases} \quad (2.8)
 \end{aligned}$$

où $z \in \mathbb{R}$, $\gamma \geq 0$, $\delta \geq 0$ (on fixe $\delta = 0$ lorsque $\gamma = 0$), $\alpha \in \mathbb{R}$, $\lambda \in \mathbb{R}$, $z_0 \geq 0$ et $K(\gamma, \delta, \alpha, \lambda, z_0)$ est une constante de normalisation. De plus, les paramètres γ , δ , α , λ et z_0 doivent remplir les conditions suivantes :

$$\mathbf{A1} : z_0 > \begin{cases} 1; & \text{si } \lambda \neq 0, \\ 0; & \text{si } \alpha \neq 0, \lambda = 0; \end{cases}$$

$$\mathbf{A2} : \alpha + \frac{\lambda}{\log z_0} + \delta \gamma z_0^\gamma \geq 0;$$

$$\mathbf{A3} : \alpha \geq 1 \text{ si } \gamma = 0;$$

$$\mathbf{A4} : \lambda > 1 \text{ si } \gamma = 0, \alpha = 1.$$

Ces différentes conditions stipulent que la densité est strictement positive et bornée (A1), qu'elle est unimodale (A2) et qu'elle est propre (A3 et A4). La densité GEP est symétrique en zéro et constante entre $-z_0$ et z_0 .

Si $f(z)$ est la densité d'une variable aléatoire Z , alors la p-crédence à droite, notée $\text{p-cred}^+(f)$ ou $\text{p-cred}^+(Z)$, est définie par Desgagné et Angers (2005) comme suit :

Définition 2. Une densité f a une p-crédence à droite $(\gamma, \delta, \alpha, \lambda)$ s'il existe une constante $K > 0$ telle que

$$\lim_{z \rightarrow \infty} \frac{f(z)}{e^{-\delta|z|^\gamma} |z|^{-\alpha} \log^{-\lambda} |z|} = K.$$

Puisque la densité GEP est symétrique par rapport à zéro, alors la définition de la p-crédence à gauche sera identique à celle de la p-crédence à droite en remplaçant $z \rightarrow \infty$ par $z \rightarrow -\infty$. La p-crédence à gauche est notée $\text{p-cred}^-(f)$ ou $\text{p-cred}^-(Z)$. La notion de p-crédence peut s'appliquer à la plupart des familles de densités connues (voir Desgagné et Angers, 2003).

La proposition suivante, de Desgagné et Angers (2007), permet de comparer les ailes de deux densités lorsque leur p-crédence (à gauche ou à droite) sont déterminées.

Proposition 1. Soit f et g , deux densités telles que $\text{p-cred}^+(f) = (\gamma, \delta, \alpha, \lambda)$ et $\text{p-cred}^+(g) = (\gamma', \delta', \alpha', \lambda')$.

i) Si $\gamma' = \gamma, \delta' = \delta, \alpha' = \alpha, \lambda' = \lambda$, alors on dit que les p-crédences à droite de f et g sont égales, notées par $(\gamma', \delta', \alpha', \lambda') = (\gamma, \delta, \alpha, \lambda)$. Leurs ailes à droite sont équivalentes, c'est-à-dire que $\lim_{z \rightarrow \infty} \frac{f(z)}{g(z)} = k$ pour une constante positive k .

ii) On dit que la p -crédence à droite de g est plus petite que celle de f , notée $(\gamma', \delta', \alpha', \lambda') < (\gamma, \delta, \alpha, \lambda)$, lorsqu'une des conditions suivantes est réalisée :

$$a) \gamma' < \gamma, b) \gamma' = \gamma, \delta' < \delta,$$

$$c) \gamma' = \gamma, \delta' = \delta, \alpha' < \alpha, d) \gamma' = \gamma, \delta' = \delta, \alpha' = \alpha, \lambda' < \lambda.$$

Ici, l'aile à droite de g domine strictement l'aile à droite de f , c'est-à-dire

$$\lim_{z \rightarrow \infty} \frac{f(z)}{g(z)} = 0.$$

Afin de comparer les ailes à gauche de deux densités, on procédera de façon similaire en utilisant la p -crédence à gauche. Ainsi, l'aile à gauche de la densité ayant une p -crédence à gauche la plus petite domine l'aile à gauche de l'autre densité.

Les propositions 2 et 3, de Desgagné et Angers (2007), permettent de simplifier les conditions C1 à C3 et les conditions des équations (2.4) à (2.7) du théorème 1.

Proposition 2. Si $p\text{-cred}^+(f) = (\gamma, \delta, \alpha, \lambda)$ et $\gamma < 1$, alors les conditions C1 à C3 sont satisfaites pour l'aile à droite de la densité f . Si $p\text{-cred}^-(f) = (\gamma, \delta, \alpha, \lambda)$ et $\gamma < 1$, alors les conditions C1 à C3 sont satisfaites pour l'aile à gauche.

Proposition 3. Si, pour des densités f_1, \dots, f_s ,

$$a) p\text{-cred}^+(f_i) = (\gamma_i, \delta_i, \alpha_i, \lambda_i), i = 1, \dots, s, \text{ et}$$

$$b) \text{ pour une même constante } c, \text{ avec } 0 < c < 1, \gamma_i \in \{0, c\}, \text{ pour } i = 1, \dots, s$$

alors

$$p\text{-cred}^+(h) = \left(\max_{i=1, \dots, s} \gamma_i, \sum_{i=1}^s \delta_i, \sum_{i=1}^s \alpha_i, \sum_{i=1}^s \lambda_i \right),$$

où la densité de h est telle que

$$h(\theta) \propto \prod_{i=1}^s f_i(\theta - u_i),$$

pour toute constante u_1, \dots, u_s . Par convention, $\delta_i = 0$ si $\gamma_i = 0$.

En utilisant les résultats des propositions 1 à 3, Desgagné et Angers (2007) formulent à nouveau les résultats a) et b) du théorème 1.

Théorème 2. Supposons que $\text{p-cred}^-(f_i) = (\gamma_i^-, \delta_i^-, \alpha_i^-, \lambda_i^-)$ et $\text{p-cred}^+(f_i) = (\gamma_i^+, \delta_i^+, \alpha_i^+, \lambda_i^+)$, tel que pour une même constante c , avec $0 < c < 1$, $\gamma_i^+, \gamma_i^- \in \{0, c\}$, pour $i = 0, \dots, n$. Pour tous nombres entiers k et m tels que $0 \leq k \leq m \leq n$ et pour toute valeur fixée $x_{j_0}, x_{j_1}, \dots, x_{j_k}$, si

$$\begin{aligned} & \left(\max_{i=k+1, \dots, m} \gamma_{j_i}^-, \sum_{i=k+1}^m \delta_{j_i}^-, \sum_{i=k+1}^m \alpha_{j_i}^-, \sum_{i=k+1}^m \lambda_{j_i}^- \right) \\ & < \left(\max_{i=0, \dots, k} \gamma_{j_i}^+, \sum_{i=0}^k \delta_{j_i}^+, \sum_{i=0}^k \alpha_{j_i}^+, \sum_{i=0}^k \lambda_{j_i}^+ \right) \text{ lorsque } k < m, \text{ et} \\ & \left(\max_{i=m+1, \dots, n} \gamma_{j_i}^+, \sum_{i=m+1}^n \delta_{j_i}^+, \sum_{i=m+1}^n \alpha_{j_i}^+, \sum_{i=m+1}^n \lambda_{j_i}^+ \right) \\ & < \left(\max_{i=0, \dots, k} \gamma_{j_i}^-, \sum_{i=0}^k \delta_{j_i}^-, \sum_{i=0}^k \alpha_{j_i}^-, \sum_{i=0}^k \lambda_{j_i}^- \right) \text{ lorsque } m < n, \end{aligned}$$

alors le résultat a) du théorème 1 est vrai.

De plus, pour toute fonction $w(\cdot)$ définie sur \mathbb{R} telle que $\mathbb{E}^{\pi(\theta|\underline{x}_k)}[|w(\theta)|] < \infty$ et $|w(\theta)|\pi(\theta|\underline{x}_k)$ est bornée supérieurement, si $\text{p-cred}^-(|w|) = (\gamma_w^-, \delta_w^-, \alpha_w^-, \lambda_w^-)$ et $\text{p-cred}^+(|w|) = (\gamma_w^+, \delta_w^+, \alpha_w^+, \lambda_w^+)$, tel que pour une même constante c , avec $0 < c < 1$, $\gamma_w^+, \gamma_w^- \in \{0, c\}$, et si

$$\begin{aligned} & \left(\max_{i=k+1, \dots, m} \gamma_{j_i}^-, \sum_{i=k+1}^m \delta_{j_i}^-, \sum_{i=k+1}^m \alpha_{j_i}^-, \sum_{i=k+1}^m \lambda_{j_i}^- \right) \\ & < \left(\max_{i=0, \dots, k} (\gamma_{j_i}^+, \gamma_w^-), \sum_{i=0}^k \delta_{j_i}^+ + \delta_w^-, \sum_{i=0}^k \alpha_{j_i}^+ + \alpha_w^-, \sum_{i=0}^k \lambda_{j_i}^+ + \lambda_w^- \right) \end{aligned}$$

lorsque $k < m$, et

$$\begin{aligned} & \left(\max_{i=m+1, \dots, n} \gamma_{j_i}^+, \sum_{i=m+1}^n \delta_{j_i}^+, \sum_{i=m+1}^n \alpha_{j_i}^+, \sum_{i=m+1}^n \lambda_{j_i}^+ \right) \\ & < \left(\max_{i=0, \dots, k} (\gamma_{j_i}^-, \gamma_w^+), \sum_{i=0}^k \delta_{j_i}^- + \delta_w^+, \sum_{i=0}^k \alpha_{j_i}^- + \alpha_w^+, \sum_{i=0}^k \lambda_{j_i}^- + \lambda_w^+ \right) \end{aligned}$$

lorsque $m < n$,

alors le résultat b) du théorème 1 est vrai.

2.6 Exemples

Dans cette section, nous considérons le jeu de données simulé à partir du modèle $y_i = 2,5x_i + \varepsilon_i$, avec $\varepsilon_i \sim N(0; 2,5^2)$ introduit à la section (1.2). Ces données sont présentées à nouveau dans le tableau 2.1.

i	x_i	y_i
1	1,1	2,4
2	1,2	0,5
3	1,5	7,6
4	3,0	10,3
5	3,7	6,4
6	3,9	9,2
7	3,9	10,7
8	5,0	15,6
9	5,7	13,6
10	7,1	18,8

Tableau 2.1 Coordonnées x et y pour les 10 points de l'ensemble de données

Puisque nous considérons le paramètre d'échelle σ connu, nous estimerons seulement le paramètre β . Le cas où σ n'est pas connu est étudié dans le prochain chapitre. Motivé par les résultats du théorème 1, nous estimons β par l'espérance *a posteriori* de β . Si on considère la notation de Desgagné et Angers (2007), et si la densité *a posteriori* de β déterminée à la section 2.3 est donnée par

$$\pi(\beta|z) = \frac{\prod_{i=0}^n f_i(z_i - \beta)}{m(z)},$$

où la densité *a priori* $\pi(\beta) = f_0(z_0 - \beta)$ et z_0 est un paramètre de position connu de la densité *a priori* de β , alors on définit l'estimateur de β par

$$b = \mathbb{E}[\beta|z] = \int_{-\infty}^{\infty} \beta \pi(\beta|z) d\beta = \frac{\int_{-\infty}^{\infty} \beta \prod_{i=0}^n f_i(z_i - \beta) d\beta}{\int_{-\infty}^{\infty} \prod_{i=0}^n f_i(z_i - \beta) d\beta}. \quad (2.9)$$

Si on considère la notation de la régression linéaire, alors

$$b = \mathbb{E}[\beta | \underline{z}, \underline{x}] = \int_{-\infty}^{\infty} \beta \pi(\beta | \underline{z}, \underline{x}) d\beta = \frac{\int_{-\infty}^{\infty} \beta \pi(\beta) \prod_{i=1}^n \frac{1}{1/x_i} f_{\varepsilon} \left(\frac{z_i - \beta}{1/x_i} \right) d\beta}{\int_{-\infty}^{\infty} \pi(\beta) \prod_{i=1}^n \frac{1}{1/x_i} f_{\varepsilon} \left(\frac{z_i - \beta}{1/x_i} \right) d\beta},$$

ou de façon équivalente,

$$b = \mathbb{E}[\beta | \underline{x}, \underline{y}] = \int_{-\infty}^{\infty} \beta \pi(\beta | \underline{x}, \underline{y}) d\beta = \frac{\int_{-\infty}^{\infty} \beta \pi(\beta) \prod_{i=1}^n f_{\varepsilon}(y_i - \beta x_i) d\beta}{\int_{-\infty}^{\infty} \pi(\beta) \prod_{i=1}^n f_{\varepsilon}(y_i - \beta x_i) d\beta}.$$

Notons qu'on peut prendre également la médiane ou le mode *a posteriori* de β comme estimateur de β , dépendant de la fonction de perte considérée.

Afin d'être le plus neutre possible quant au choix de la distribution *a priori* de β , nous avons choisi $\pi(\beta) = f_0(z_0 - \beta) \equiv 1$. La distribution *a priori* est donc une distribution impropre et non informative. Les ailes d'une telle densité ne tendent pas vers zéro de telle sorte que l'information *a priori* ne sera jamais en conflit avec les autres sources d'information. On remarque qu'avec ce choix, la densité *a posteriori* est proportionnelle à la vraisemblance. Si on considérait en plus le mode *a posteriori* de β comme estimateur de Bayes de β , alors cela reviendrait à trouver l'estimateur du maximum de vraisemblance. Par exemple, si on suppose que la variable aléatoire ε_i est distribuée selon une loi $N(0; \sigma^2)$, pour $i = 1, \dots, n$, alors la variable aléatoire $Z_i = Y_i/X_i = \beta + \varepsilon_i/X_i$, sachant β et X_i , est distribuée selon une loi $N(\beta; (\sigma/x_i)^2)$. Si on considère la variable $\beta | \underline{z}, \underline{x}$ alors

par l'équation (2.3) et avec $\pi(\beta) = f_0(z_0 - \beta) \equiv 1$, on a

$$\begin{aligned}
\pi(\beta|\underline{z}, \underline{x}) &\propto \prod_{i=1}^n \frac{1}{1/x_i} f_\varepsilon\left(\frac{z_i - \beta}{1/x_i}\right) \\
&\propto \prod_{i=1}^n \exp\left\{\frac{-1}{2\sigma^2} \left(\frac{z_i - \beta}{1/x_i}\right)^2\right\} \\
&\propto \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n \left(\frac{z_i - \beta}{1/x_i}\right)^2\right\} \\
&\propto \exp\left\{\frac{-1}{2\sigma^2} \left(\sum_{i=1}^n (x_i z_i)^2 + \beta^2 \sum_{i=1}^n x_i^2 - 2\beta \sum_{i=1}^n (x_i^2 z_i)\right)\right\} \\
&\propto \exp\left\{\frac{-1}{2\sigma^2} \left(\beta^2 \sum_{i=1}^n x_i^2 - 2\beta \sum_{i=1}^n (x_i^2 z_i)\right)\right\} \\
&\propto \exp\left\{\frac{-\sum_{i=1}^n x_i^2}{2\sigma^2} \left(\beta^2 - \frac{2\beta \sum_{i=1}^n (x_i^2 z_i)}{\sum_{i=1}^n x_i^2}\right)\right\} \\
&\propto \exp\left\{\frac{-\sum_{i=1}^n x_i^2}{2\sigma^2} \left(\beta^2 - \frac{2\beta \sum_{i=1}^n (x_i^2 z_i)}{\sum_{i=1}^n x_i^2}\right) \right. \\
&\quad \left. + \left(\frac{\sum_{i=1}^n (x_i^2 z_i)}{\sum_{i=1}^n x_i^2}\right)^2 - \left(\frac{\sum_{i=1}^n (x_i^2 z_i)}{\sum_{i=1}^n x_i^2}\right)^2\right\} \\
&\propto \exp\left\{\frac{-\sum_{i=1}^n x_i^2}{2\sigma^2} \left(\beta - \frac{\sum_{i=1}^n (x_i^2 z_i)}{\sum_{i=1}^n x_i^2}\right)^2\right\}.
\end{aligned}$$

C'est donc dire que si on suppose que la variable aléatoire ε_i de l'équation (2.1) est distribuée selon une loi $N(0; \sigma^2)$ pour un σ^2 connu, alors la variable aléatoire $\beta|\underline{z}, \underline{x}$ est distribuée selon une loi $N\left(\frac{\sum_{i=1}^n x_i^2 z_i}{\sum_{i=1}^n x_i^2}; \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right)$. L'espérance, la médiane et le mode d'une variable aléatoire normale étant tous égaux, on voit bien que l'espérance *a posteriori* de β n'est nul autre que l'estimateur de vraisemblance maximal de β et est égal à $\frac{\sum_{i=1}^n x_i^2 z_i}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$, soit l'estimateur de β vu à la section 1.2. Cependant, la loi $N(0; \sigma^2)$ ne permet pas d'obtenir des estimateurs robustes comme nous l'avons constaté au chapitre 1. Cela s'explique par les résultats théoriques de robustesse, car les ailes de la densité de la loi normale ne sont pas suffisamment relevées, comme nous allons le voir dans cette section. Dans le but d'obtenir un estimateur bayésien de β robuste, nous considérons plutôt les lois de Student et GEP de type V afin de décrire le comportement de la variable aléatoire ε_i . Notons que ε_i et Z_i ne diffèrent que par un paramètre de position et échelle, car $\varepsilon_i = \frac{Z_i - \beta}{1/X_i}$. La densité GEP de type V est donnée par l'équation (2.8) avec

les paramètres $\gamma = 0$, $\delta = 0$, $\alpha = 1$ et $\lambda > 1$ (Desgagné et Angers, 2005). Nous avons fait le choix de ces deux lois puisqu'elles possèdent toutes deux des ailes plus relevées que celles de la loi normale. Nous allons le démontrer en comparant la p-crédence à droite de ces trois lois symétriques, en utilisant la définition 2 et la proposition 2 de la section 2.5.2. Tout d'abord, si une variable aléatoire Z_1 est distribuée selon une loi $N(0; \sigma^2)$ avec densité $f_{Z_1}(z) \propto \exp\{\frac{-z^2}{2\sigma^2}\}$, alors $p\text{-cred}^+(Z_1)$ correspond au choix des paramètres $(\gamma, \delta, \alpha, \lambda)$ qui satisfont

$$\lim_{z \rightarrow \infty} \frac{f(z)}{e^{-\delta|z|^\gamma} |z|^{-\alpha} \log^{-\lambda} |z|} = K,$$

pour un certain $K > 0$. Ainsi, si on choisit $\gamma = 2$, $\delta = \frac{1}{2\sigma^2}$, $\alpha = 0$ et $\lambda = 0$, alors

$$\begin{aligned} \lim_{z \rightarrow \infty} \frac{f(z)}{e^{-\delta z^\gamma} z^{-\alpha} \log^{-\lambda}(z)} &= \lim_{z \rightarrow \infty} \frac{K \exp\{\frac{-z^2}{2\sigma^2}\}}{e^{-\delta z^\gamma} z^{-\alpha} \log^{-\lambda}(z)} \\ &= \lim_{z \rightarrow \infty} \frac{K \exp\{\frac{-z^2}{2\sigma^2}\}}{\exp\{\frac{-z^2}{2\sigma^2}\}} \\ &= K, \end{aligned}$$

où K est la constante de normalisation de la loi normale. On obtient donc, $p\text{-cred}^+(Z_1) = (2, \frac{1}{2\sigma^2}, 0, 0)$. Maintenant, soit une variable aléatoire Z_2 qui est distribuée selon une loi de Student avec ν degrés de liberté, pour $\nu > 0$, et dont la densité est $f_{Z_2}(z) \propto \left(1 + \frac{z^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}$. Si on choisit $\gamma = 0$, $\delta = 0$, $\alpha = \nu + 1$ et $\lambda = 0$, alors

$$\begin{aligned} \lim_{z \rightarrow \infty} \frac{f(z)}{e^{-\delta z^\gamma} z^{-\alpha} \log^{-\lambda}(z)} &= \lim_{z \rightarrow \infty} \frac{K (\nu + z^2)^{-\frac{(\nu+1)}{2}}}{e^{-\delta z^\gamma} z^{-\alpha} \log^{-\lambda}(z)} \\ &= \lim_{z \rightarrow \infty} \frac{K (\nu + z^2)^{-\frac{(\nu+1)}{2}}}{e^{-\delta z^\gamma} (z^2)^{-\frac{\alpha}{2}} \log^{-\lambda}(z)} \\ &= \lim_{z \rightarrow \infty} \frac{K (\nu + z^2)^{-\frac{(\nu+1)}{2}}}{(z^2)^{-\frac{\alpha}{2}}} \\ &= \lim_{z \rightarrow \infty} K \left(\frac{\nu + z^2}{z^2} \right)^{-\frac{(\nu+1)}{2}} \\ &= \lim_{z \rightarrow \infty} K \left(\frac{\nu}{z^2} + 1 \right)^{-\frac{(\nu+1)}{2}} \\ &= K, \end{aligned}$$

où K est la constante de normalisation de la loi Student. On obtient alors $\text{p-cred}^+(Z_2) = (0, 0, \nu + 1, 0)$. Enfin, si une variable aléatoire Z_3 est distribuée selon une GEP de type V, il est trivial que $\text{p-cred}^+(Z_3) = (0, 0, 1, \lambda)$.

Afin de comparer les ailes à droite des densités deux à deux, nous utilisons la proposition 2. Si on note $\text{p-cred}^+(Z_1) = (\gamma, \delta, \alpha, \lambda)$, $\text{p-cred}^+(Z_2) = (\gamma', \delta', \alpha', \lambda')$ et $\text{p-cred}^+(Z_3) = (\gamma'', \delta'', \alpha'', \lambda'')$, alors $\text{p-cred}^+(Z_2) < \text{p-cred}^+(Z_1)$, car $\gamma' < \gamma$, où $\gamma' = 0$ et $\gamma = 2$. L'aile à droite de la densité de Student domine strictement l'aile à droite de la densité normale. Par symétrie, la densité de Student a des ailes plus relevées que celles de la loi $N(0; \sigma^2)$. Aussi, $\text{p-cred}^+(Z_3) < \text{p-cred}^+(Z_2)$, car $\gamma'' = \gamma'$, $\delta'' = \delta'$ et $\alpha'' < \alpha'$, où $\alpha'' = 1$ et $\alpha' = \nu + 1$ avec $\nu > 0$. L'aile à droite de la densité GEP de type V domine strictement l'aile à droite de la densité Student. Par symétrie, la densité GEP de type V a donc des ailes plus relevées que celles de la Student, et par conséquent ses ailes sont aussi plus relevées que celles de la normale. La densité Student est appelée densité à ailes relevées tandis que la densité GEP de type V est appelée densité à ailes super relevées (Desgagné et Angers, 2005).

2.6.1 Student

À l'origine, les données ont été simulées en posant la variable aléatoire ε_i i.i.d $N\left(0; (2,5)^2\right)$. Cela revient à dire que $\frac{\varepsilon_i}{2,5}$ i.i.d $N(0; 1)$, soit avec un paramètre d'échelle de 2,5. Puisque nous supposons le paramètre d'échelle σ connu, nous lui assignons cette valeur de 2,5. Dans cette section, nous supposons plutôt que la variable aléatoire ε_i est distribuée selon une loi F avec paramètre d'échelle σ connu et égal à 2,5, où F est une loi de Student avec 4 degrés de liberté et paramètre d'échelle connu τ_0 . La densité des ε_i est

$$f_{\varepsilon}(\varepsilon_i) \propto \left(1 + \frac{1}{4} \left(\frac{\varepsilon_i}{2,5\tau_0}\right)^2\right)^{-5/2}.$$

Afin d'influencer le moins possible l'inférence en cas de non-conflit, nous fixons le paramètre τ_0 de telle sorte que la variable aléatoire $\frac{\varepsilon_i}{2,5}$ ait une densité F (soit une loi de Student avec 4 degrés de liberté et paramètre d'échelle τ_0) ayant une distance inter-

quartile égale à celle d'une $N(0; 1)$. Considérons les distances interquartiles

$$\begin{aligned}
 d_N &= Q_N(3) - Q_N(1) \\
 &= 0,6745 - (-0,6745) \\
 &= 1,349, \\
 d_T &= Q_T(3) - Q_T(1) \\
 &= 0,7407 - (-0,7407) \\
 &= 1,4814,
 \end{aligned}$$

où $Q_N(3)$ et $Q_N(1)$ sont les troisième et premier quartiles de la loi normale centrée réduite et $Q_T(3)$ et $Q_T(1)$ sont les troisième et premier quartiles de la loi de Student avec 4 degrés de liberté. Alors τ_0 est tel que

$$\begin{aligned}
 d_T \tau_0 &= d_N \\
 \Leftrightarrow \tau_0 &= \frac{d_N}{d_T} \\
 \Leftrightarrow \tau_0 &= 0,9106.
 \end{aligned}$$

On voit sur la figure 2.1 qu'avec ce choix, la distribution de la variable aléatoire $\frac{\varepsilon_i}{2,5}$ a une dispersion semblable à celle d'une normale centrée réduite, sauf pour les ailes. Puisque nous considérons plutôt la variable aléatoire $Z_i = \frac{Y_i}{X_i}$, alors par l'équation (2.2), la densité de cette variable évaluée au point $z_i = \frac{y_i}{x_i}$, est donnée par

$$\frac{1}{1/x_i} f_\varepsilon \left(\frac{z_i - \beta}{1/x_i} \right) \propto \frac{1}{1/x_i} \left(1 + \frac{1}{4} \left(\frac{z_i - \beta}{2,2765/x_i} \right)^2 \right)^{-5/2},$$

pour $i = 1, \dots, 10$ et où $\sigma\tau_0 = 2,2765$. Nous noterons la loi de $Z_i|X_i, \beta$ par $t_4(\beta; \frac{2,2765}{x_i})$, soit une Student avec 4 degrés de liberté, avec paramètre de position β et paramètre d'échelle $\frac{2,2765}{x_i}$.

Considérons d'abord le cas où le point P_{10} est aberrant en ordonnée. Ce cas est représenté sur les graphiques (a) et (b) de la figure 2.2. Les points $P_{10}^{(I)} = (7,1; 68,8)$ et $P_{10}^{(II)} = (7,1; -31,2)$ respectivement, semblent avoir très peu ou pas d'influence sur l'orientation

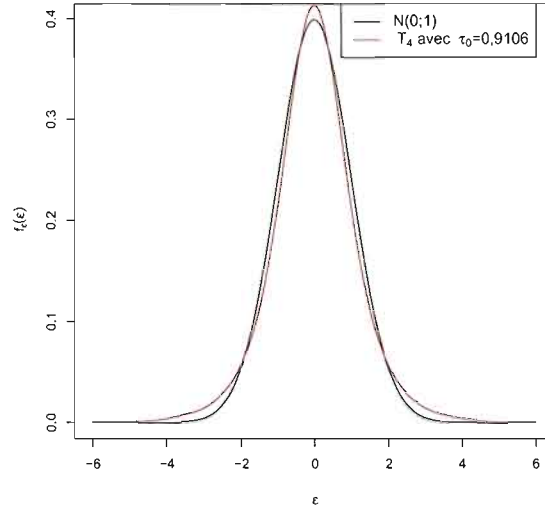


Figure 2.1 Densités $N(0;1)$ et F où F est une loi de Student avec 4 degrés de liberté et paramètre d'échelle $\tau_0 = 0,9106$.

de la droite. En effet, cette dernière emprunte la trajectoire linéaire décrite par les neuf autres points non aberrants. Pour chacun des graphiques, la pente de la droite est estimée à $b_I = 2,6549$ et $b_{II} = 2,5672$. Lorsque y_{10} tend vers plus ou moins l'infini, l'estimé de la pente converge vers la valeur $b_{Student} = 2,6103$ tel qu'illustré sur la figure 2.3. Cette valeur, correspond à l'estimé de β lorsque le point P_{10} est retiré de l'ensemble de données. La méthode est donc robuste à une valeur aberrante en ordonnée.

Afin de mieux comprendre, analysons le cas où y_{10} tend vers l'infini. Considérons les graphiques (a), (b) et (c) de la figure 2.4, représentant le positionnement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ selon la valeur de la coordonnée y_{10} . Notons que X_1, \dots, X_{10} sont considérées implicitement connues pour alléger la notation. Aussi, la densité *a posteriori* $\pi(\beta|Z_{10})$, qui est proportionnelle à la densité $g_{Z_{10}}(z_{10} - \beta|\beta, x_{10})$, est centrée en $\beta = z_{10} = y_{10}/x_{10}$. Dans la première figure, $y_{10} = 18,8$ ($z_{10} = y_{10}/x_{10} = 2,6479$), ce qui correspond au jeu de données original du tableau 2.1.

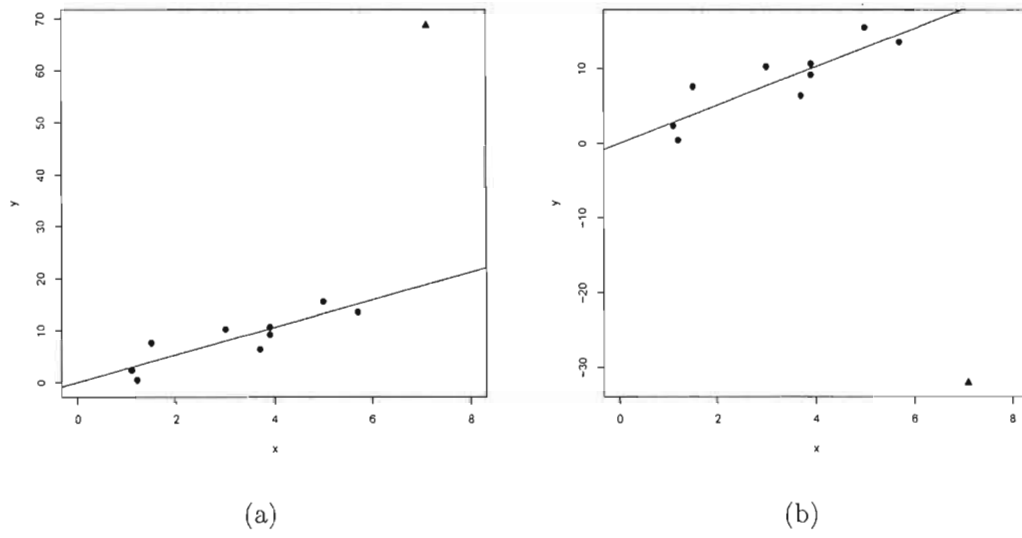


Figure 2.2 Estimé bayésien de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7, 1; 68, 8)$, (b) lorsque le point $P_{10}^{(II)} = (7, 1; -31, 2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. Les variables aléatoires Z_i suivent une loi $t_4(\beta; \frac{2,2765}{x_i})$.

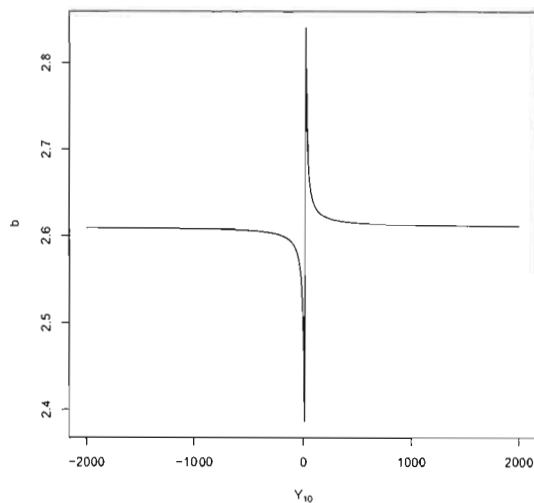


Figure 2.3 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} .

L'estimé de β est égal à 2,6229 alors qu'il n'y a pas d'information conflictuelle, c'est-à-dire qu'il n'y a pas de point aberrant. Le point P_{10} n'est pas écarté par la méthode d'estimation, mais puisqu'il se retrouve dans la trajectoire linéaire formée par les neuf autres points, son influence sur l'estimé de β paraît négligeable ($b = 2,6229$ alors que $b_{Student} = 2,6103$). Cela se traduit par le fait que les densités *a posteriori* $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ ont une importante zone commune due au fait qu'elles sont superposées. La densité $\pi(\beta|Z_1, \dots, Z_{10})$ se trouve au même endroit que les deux densités $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$, comme on peut l'observer sur le graphique. La densité *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$ est donc calculée en fonction du compromis le plus réaliste entre les différentes sources d'information z_1, \dots, z_{10} . Rappelons que la densité *a priori* de β est $\pi(\beta) \equiv 1$, d'où l'absence d'un paramètre de position z_0 pour cette densité.

Sur le graphique (b) de la figure 2.4, le point P_{10} se retrouve maintenant aux coordonnées (7,1; 28,8), ce qui donne $z_{10} = 4,0563$. À nouveau, il n'y a pas d'information conflictuelle. Cependant, il est intéressant de constater comment se comporte la den-

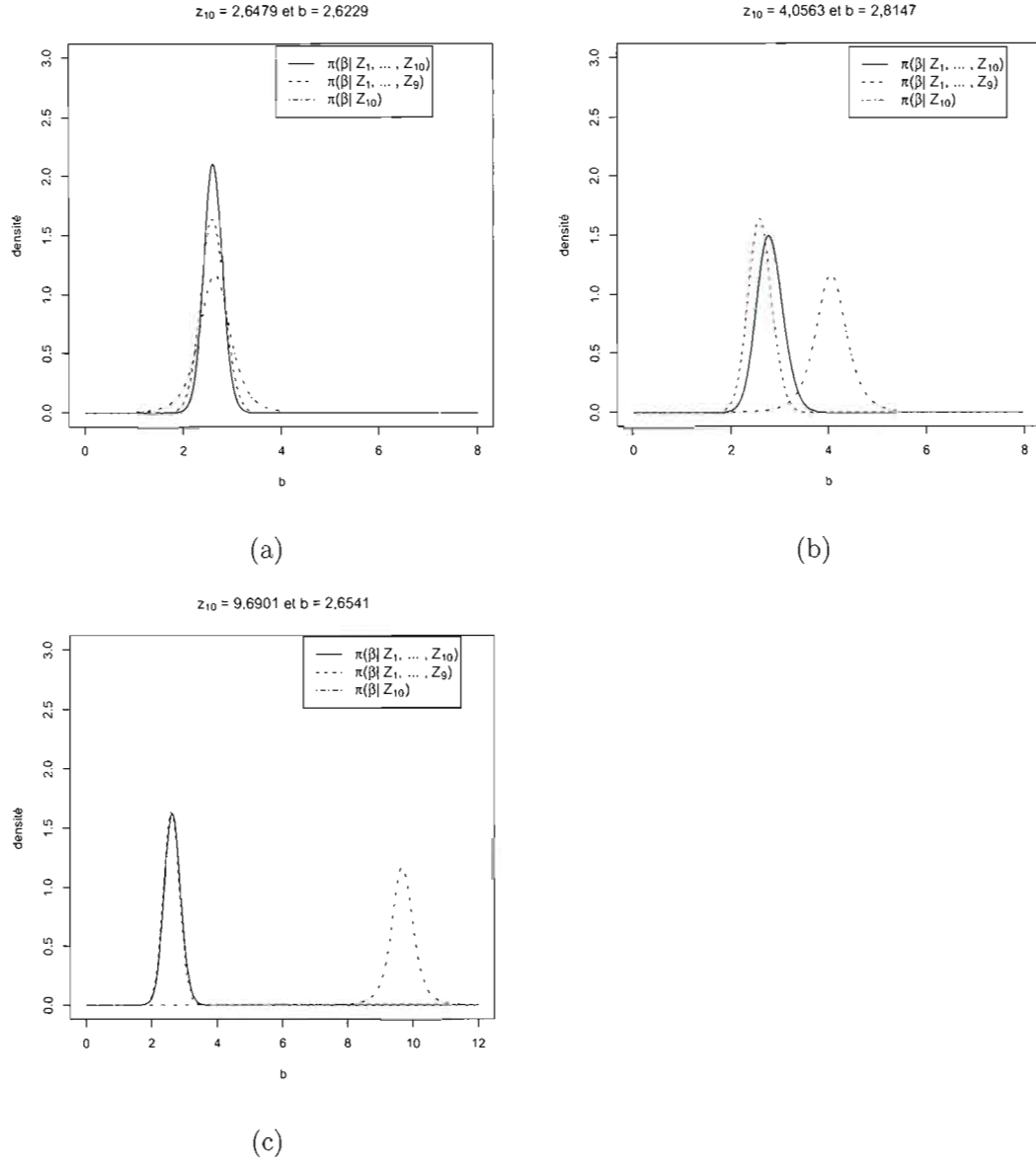


Figure 2.4 Comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque (a) $y_{10} = 18,8$, (b) $y_{10} = 28,8$ et (c) $y_{10} = 68,8$.

sité *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque les densités $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ ne sont plus du tout superposées comme dans le cas précédent. On voit que les densités *a posteriori* $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ se chevauchent maintenant dans les ailes et que leur zone commune a diminué par rapport au cas précédent. L'influence de la valeur $y_{10} = 28,8$ ($z_{10} = 4,0563$) est à son maximum et se traduit par une estimation de β de 2,8147.

Enfin, sur le graphique (c) de la figure 2.4, nous illustrons le cas où le point $P_{10}^{(f)} = (7,1 ; 68,8)$ est aberrant, soit $z_{10} = 9,6901$. En présence d'information conflictuelle, on remarque que la zone commune entre les densités *a posteriori* $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ est maintenant négligeable. L'importance accordée au point P_{10} dans le calcul de l'estimé de β est beaucoup plus faible que dans le cas précédent. Cela se traduit par le fait que $b_I = 2,6549$ est beaucoup plus près de $b_{Student} = 2,6103$ que l'estimé $b = 2,8147$ calculé au cas correspondant à la figure 2.4 (b). La densité *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$ est revenue vers la densité *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$. Lorsque y_{10} (z_{10}) tend vers l'infini, la méthode d'estimation écarte graduellement l'information conflictuelle (le point aberrant P_{10}) au profit des neuf autres points. La densité $\pi(\beta|Z_1, \dots, Z_{10})$ se comporte de plus en plus comme la densité $\pi(\beta|Z_1, \dots, Z_9)$ et à la limite, les deux densités seront identiques. Ainsi, l'estimé de la pente calculé à partir de la densité *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$ tend vers l'estimé de la pente calculé à partir de $\pi(\beta|Z_1, \dots, Z_9)$ à mesure que y_{10} (z_{10}) tend vers plus l'infini, tel que stipulé par le théorème 1.

Considérons maintenant le cas où le point P_{10} est aberrant en abscisse. Rappelons que la variable aléatoire $Z_i = \frac{Y_i}{X_i}$ où X_i est connue. Alors,

$$\begin{aligned} \text{Var}[Z_{10}|x_{10}, \beta] &= \text{Var}[Y_{10}/X_{10}|x_{10}, \beta] \\ &= \frac{1}{x_{10}^2} \text{Var}[Y_{10}|x_{10}, \beta] \\ &= \frac{1}{x_{10}^2} \text{Var}[\varepsilon_{10}]. \end{aligned}$$

La variance de la variable aléatoire Z_{10} est d'autant petite que x_{10} tend vers plus ou moins l'infini. La théorie sur la robustesse vue à la section 2.5 suppose que les

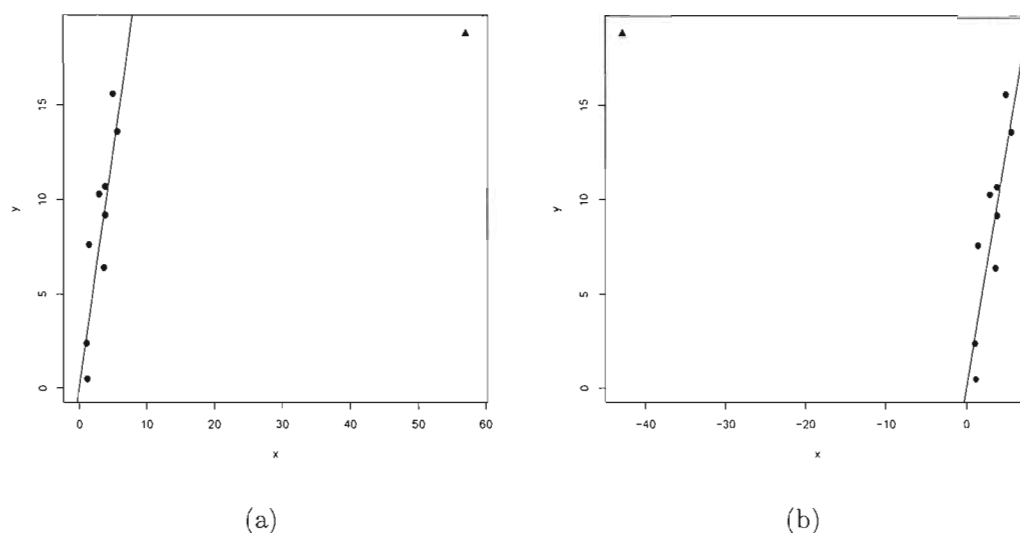


Figure 2.5 Estimé bayésien de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57, 1; 18, 8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42, 9; 18, 8)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. Les variables aléatoires Z_i suivent une loi $t_4(\beta; \frac{2,2765}{x_i})$.

variables aléatoires peuvent provenir de lois différentes n'ayant pas nécessairement la même variance. Cependant, même si la réalisation de certaines de ces variables tend vers plus ou moins l'infini, leur variance respective est considérée fixe, contrairement à ce qui est observé ici. Puisque notre analyse des méthodes vues précédemment portait aussi sur la robustesse par rapport à une valeur aberrante en abscisse, nous allons tout de même explorer le cas où la coordonnée x_{10} tend vers plus ou moins l'infini, malgré le fait que les résultats sur la robustesse ne s'appliquent pas.

Les graphiques (a) et (b) de la figure 2.5 illustrent le nuage de points lorsque le point aberrant est localisé aux coordonnées $P_{10}^{(III)} = (57, 1; 18, 8)$ et $P_{10}^{(IV)} = (-42, 9; 18, 8)$, respectivement. L'estimé de β est donné par $b_{III} = 2,4624$ et $b_{IV} = 2,5049$. Ces deux valeurs sous-estiment la valeur de b lorsque le point P_{10} est retiré de l'ensemble de données, c'est-à-dire $b_{Student} = 2,6103$. En analysant seulement la figure 2.5, la méthode

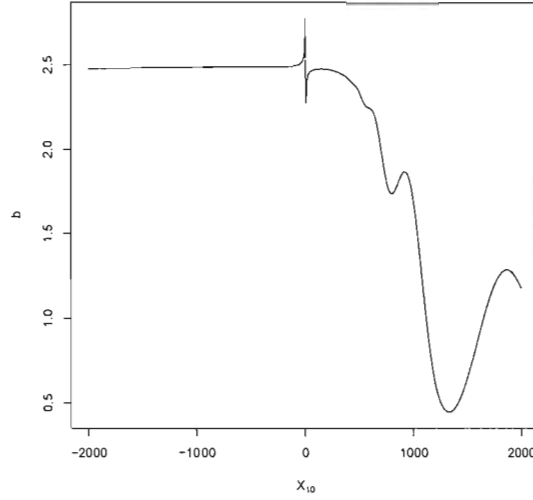


Figure 2.6 Comportement de l'estimé b pour différentes valeurs de la coordonnée x_{10} .

d'estimation bayésienne lorsque les variables Z_i sont distribuées selon une $t_4(\beta; \frac{2,2765}{x_i})$, semble être robuste. Cependant, la figure 2.6 démontre plutôt que lorsque x_{10} tend vers l'infini (z_{10} tend vers zéro), l'estimé de β converge vers 0. Le phénomène est aussi vrai lorsque x_{10} tend vers moins l'infini, mais l'estimé b tend vers zéro à une vitesse beaucoup plus lente.

Afin de mieux comprendre, considérons les graphiques (a), (b) (c) et (d) de la figure 2.7. Ces graphiques représentent le positionnement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ pour différentes valeurs de x_{10} lorsque nous considérons le cas où x_{10} tend vers l'infini (z_{10} tend vers zéro). Tout d'abord, nous avons omis le cas où la valeur de x_{10} égale 7,1 puisqu'il correspond exactement au cas représenté dans le graphique (a) de la figure 2.4 discuté avant. Maintenant, dans le graphique (a) de la figure 2.7, le point P_{10} a comme coordonnées (17,1 ; 18,8), soit $z_{10} = 1,0994$. Aucune information n'est en conflit. À nouveau, comme pour le graphique (b) de la figure 2.4, l'intérêt est d'analyser le comportement de la densité *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$

lorsque la zone commune des densités $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ est restreinte. Le changement de la coordonnée x_{10} du point P_{10} , passant de 7,1 dans le graphique (a) de la figure 2.4 à 17,1 dans le graphique (a) de la figure 2.7 a eu pour effet de diminuer la variance de la variable Z_{10} (et donc de $\beta|z_{10}$) et de déplacer la densité $\pi(\beta|Z_{10})$ vers la gauche, tel qu'observé dans le graphique (a) de la figure 2.7. Cela a entraîné la réduction de la zone commune entre les densités $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$, qui a eu pour conséquence d'augmenter la dispersion de la densité $\pi(\beta|Z_1, \dots, Z_{10})$ tout en déplaçant cette densité vers la gauche, mais de façon plus modeste. L'estimé de β se voit être affecté par l'influence non négligeable du point P_{10} , chutant de 2,6229 dans la figure 2.4 (a) à 2,351 dans la figure 2.7 (a).

Sur le graphique (b) de la figure 2.7, nous avons considéré le point aberrant $P_{10}^{(III)}$ dont les coordonnées sont (57,1 ; 18,8), ce qui donne $z_{10} = 0,3292$. L'estimé de β est 2,4624. On voit que la densité $\pi(\beta|Z_{10})$ est trop éloignée de la densité $\pi(\beta|Z_1, \dots, Z_9)$ de telle sorte que la zone commune entre les deux densités semble suffisamment petite pour créer une situation de conflit. La dispersion de $\pi(\beta|Z_{10})$ est beaucoup moins grande due au fait que plus x_{10} tend vers l'infini, plus la variance de la variable aléatoire Z_{10} tend vers 0. Cependant, malgré que les résultats théoriques de la section 2.5 n'ont pas été développés dans le cas particulier d'une variance qui varie en fonction de l'information, la méthode d'estimation bayésienne produit néanmoins un estimé de β qui semble être peu influencé par la valeur aberrante en x . Tout comme une méthode robuste, la densité $\pi(\beta|Z_1, \dots, Z_{10})$ se rapproche de $\pi(\beta|Z_1, \dots, Z_9)$, indiquant que la méthode d'estimation donne moins de poids au point P_{10} et privilégie les points P_1, \dots, P_9 dans le calcul de b . Cependant, comme on l'a vu, l'estimé de β tend vers 0 à mesure que x_{10} tend vers plus ou moins l'infini. Cette situation devient plus claire dans les deux prochains cas.

Sur le graphique (c) de la figure 2.7, la coordonnée x_{10} égale 407,1, soit $z_{10} = 0,0462$. On voit que la densité $\pi(\beta|Z_{10})$ s'est déplacée vers la gauche et que sa variance a encore diminué. Quant à elle, la densité $\pi(\beta|Z_1, \dots, Z_{10})$ s'est à peine déplacée vers la gauche, mais on est maintenant en mesure de constater que cette densité est bimodale. En effet, on remarque qu'un mode se retrouve près de la densité $\pi(\beta|Z_{10})$, alors que le

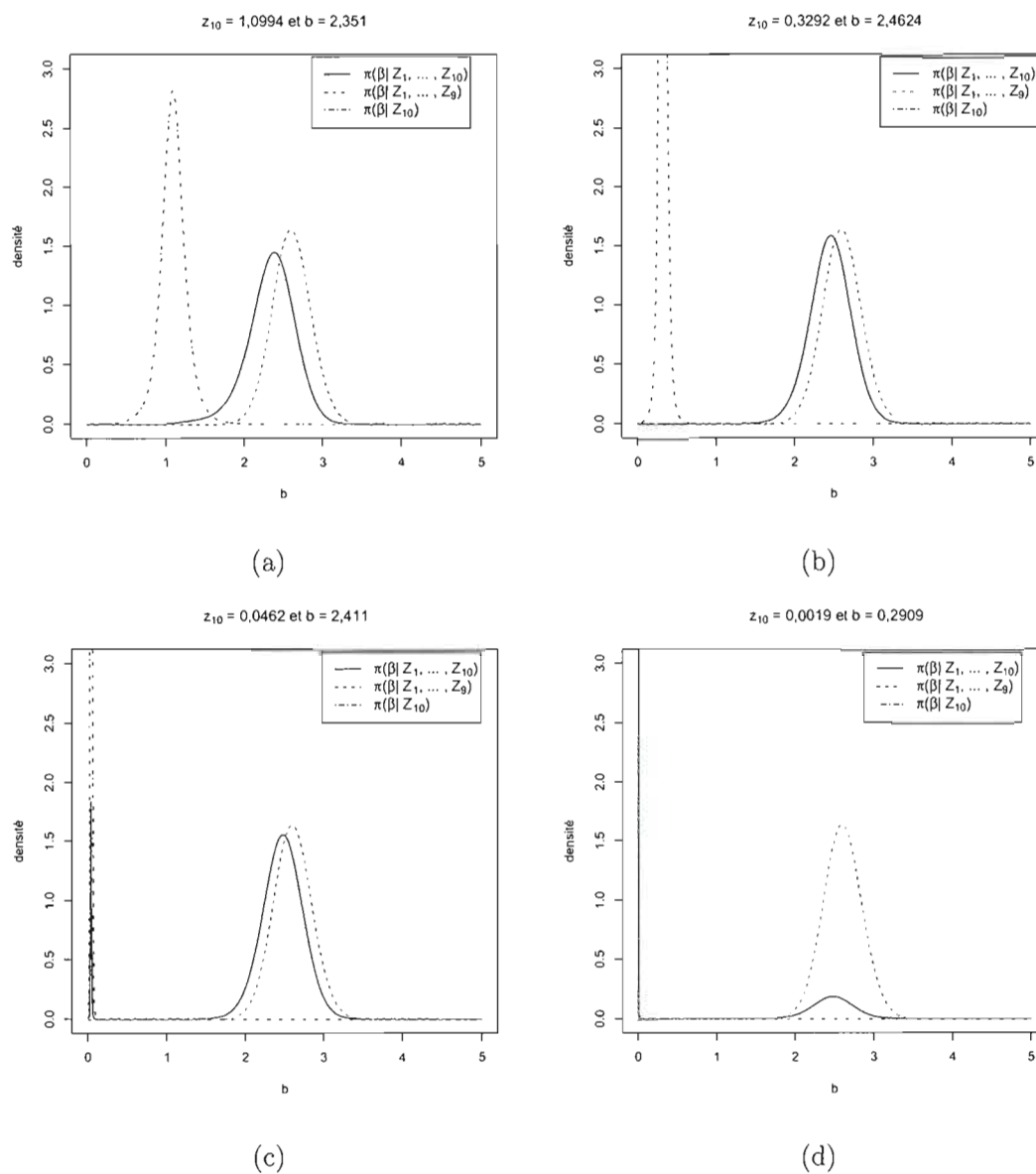


Figure 2.7 Comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque (a) $x_{10} = 17,1$, (b) $x_{10} = 57,1$, (c) $x_{10} = 407,1$ et (d) $x_{10} = 10007,1$.

deuxième mode et la densité $\pi(\beta|Z_1, \dots, Z_9)$ sont pratiquement superposés. Le point P_{10} de coordonnées $(407,1; 18,8)$, bien qu'ayant une influence limitée, n'est pas écartée par la méthode d'estimation. La densité $\pi(\beta|Z_1, \dots, Z_{10})$ est calculée en faisant un compromis qui pourrait sembler acceptable entre les données et la valeurs aberrante compte tenu du fait que l'estimé de β vaut 2,411. Cependant, comme le laisse sous-entendre la figure 2.6, l'influence du point P_{10} augmente à mesure que x_{10} tend vers l'infini puisque l'estimé de β tend vers zéro. Comme on aimerait qu'une méthode robuste accorde de moins en moins d'importance à un point aussi éloigné que le point $P_{10} = (407; 18,8)$, il est clair que la méthode bayésienne, dont la variable Z est distribuée selon une loi $t_4(\beta; \frac{2,2765}{x_i})$, n'est pas robuste à une valeur aberrante en x . Mais bien qu'elle ne soit pas robuste, cette méthode limite tout de même l'influence de certains points pouvant être considérés aberrant, comme par exemple le point de coordonnée $(-100; 18,8)$, tel qu'illustré sur la figure 2.6. L'explication de la lente convergence de l'estimé b vers 0, donnant ainsi l'impression d'une méthode robuste, peut s'expliquer par le fait qu'une valeur aberrante en abscisse, pour notre modèle $y_i = \beta x_i + \varepsilon_i$, borne les résidus qui, contrairement au cas où il y a une valeur aberrante en ordonnée, ne peuvent tendre vers l'infini. En effet, lorsque y_{10} tend vers plus ou moins l'infini, l'estimé de la pente d'une méthode non robuste sera tirée par le point aberrant. À la limite, la pente de la droite sera infinie et les résidus, en valeurs absolues, (la distance verticale entre les points et la droite) seront aussi infinis. Mais lorsqu'on considère plutôt qu'un point en abscisse tend vers plus ou moins l'infini, alors la pente de la droite, estimée d'une méthode non robuste, tend vers 0. Cela entraîne que les résidus sont bornées puisqu'ils tendent vers une limite finie. À mesure que le point aberrant s'écarte de la tendance linéaire en abscisse, les résidus découlant de l'estimation de β tendent vers leur limite respective mais à une vitesse beaucoup trop lente pour que la méthode écarte les points P_1, \dots, P_9 du calcul de b . Par exemple, sur la figure 2.8 nous avons supposé que la droite passait par l'origine et le point de coordonnée $(0,1; 18,8)$ (la droite rouge) et par l'origine et le point de coordonnée $(-100; 18,8)$ (la droite bleue). En tenant compte des dix points, l'estimé de β lorsque le point $P_{10} = (0,1; 18,8)$ est 2,6119 alors que l'estimé de β lorsque le point $P_{10} = (-100; 18,8)$ est 2,4942. Rappelons que l'estimé de β lorsque le point P_{10} est

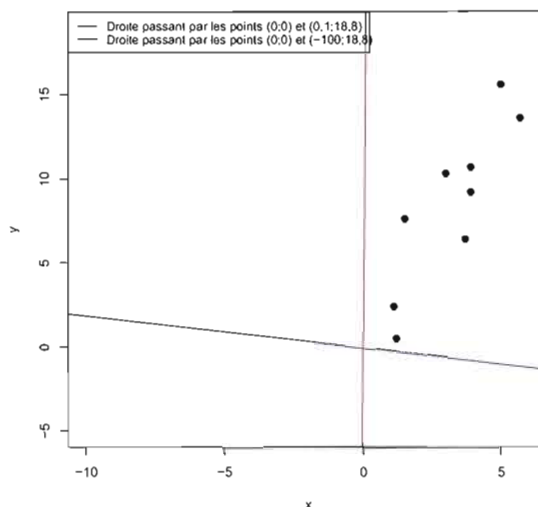


Figure 2.8 Exemple de droites si le modèle rejette les points P_1, \dots, P_9 .

retiré des données est 2,6103. On voit donc que la méthode écarte dans une proportion plus grande le point $P_{10} = (0,1; 18,8)$, qui est pourtant beaucoup plus près des données originales que ne l'est le point $P_{10} = (-100; 18,8)$. Cela est dû au fait que, un peu comme pour une valeur aberrante en y , les valeurs résiduelles produites par la droite rouge de la figure 2.8 sont beaucoup trop grandes. Le modèle écarte alors le point $P_{10} = (0,1; 18,8)$ comme un point faisant partie de l'ensemble. L'estimé b est donc semblable à celui fait à partir des neuf points P_1, \dots, P_9 . Par contre, lorsque nous considérons le point $P_{10} = (-100; 18,8)$, les résidus découlant de la droite bleue sont beaucoup plus réalistes. Le point $P_{10} = (-100; 18,8)$ n'est donc pas totalement écarté par la méthode comme ce fut le cas pour le point $P_{10} = (0,1; 18,8)$.

Analysons maintenant un cas où le conflit entre les sources d'information est encore plus flagrant. Lorsque le point aberrant P_{10} a comme coordonnée $x_{10} = 10007,1$ ($z_{10} = 0,0019$), les densités *a posteriori* se comportent tel qu'illustré sur le graphique (d) de la figure 2.7. On voit que la densité $\pi(\beta|Z_1, \dots, Z_{10})$ est toujours bimodale, mais cette fois-

ci, le mode global se retrouve du côté de la densité $\pi(\beta|Z_{10})$, signifiant que la méthode n'est pas robuste. Sans avoir totalement écarté les neuf points P_1, \dots, P_9 , la méthode leur accorde de moins en moins d'importance dans l'estimation de β au profit du point P_{10} .

Bien que la méthode d'estimation bayésienne, dont les variables Z_i sont distribuées selon une loi $t_4(\beta; \frac{2,2765}{x_i})$, produise un estimé de β robuste à une valeur aberrante en y , cela n'est pas le cas lorsqu'il y a présence d'une valeur aberrante en abscisse. Dans cette section, nous avons supposé que les variables aléatoires ε_i , pour $i = 1, \dots, n$, sont distribuées selon une loi de Student avec $\nu = 4$ degrés de liberté et avec paramètre d'échelle connu τ_0 . Le choix de ν a été motivé par le fait qu'une variable aléatoire étant distribuée selon une Student avec ν degrés de liberté, est à peu près distribuée selon une loi normale centrée réduite, lorsque ν tend vers l'infini, c'est-à-dire une densité ayant des ailes moins relevées. Ainsi, plus on choisira une valeur de ν grande par rapport à $\nu = 4$, moins on s'attend à obtenir un estimateur de β robuste, par rapport à l'estimateur de β lorsque $\nu = 4$.

2.6.2 GEP de type V

Dans cette section, nous supposons que la variable aléatoire ε_i est distribuée selon une loi F avec paramètre d'échelle σ connu égal à 2,5, où F est une loi GEP de type V avec paramètres $\gamma = \delta = 0$, $\alpha = 1$, $\lambda = 4$, $z_0 = 2$ et paramètre d'échelle connu τ_0 . La densité de ε_i est

$$f_{\varepsilon}(\varepsilon_i) \propto \left(\left| \frac{\varepsilon_i}{2,5\tau_0} \right| + 2 \right)^{-1} \left(\log \left(\left| \frac{\varepsilon_i}{2,5\tau_0} \right| + 2 \right) \right)^{-4}.$$

Notons que nous utilisons une forme modifiée de la densité GEP (Desgagné et Angers 2003) qui n'a pas de partie uniforme au centre (voir figure 2.9).

Le paramètre d'échelle τ_0 est déterminé de telle sorte que la variable aléatoire $\frac{\varepsilon_i}{2,5}$ ait une densité F ayant une distance interquartile égale à celle d'une $N(0; 1)$. Considérons

les distances interquartiles

$$\begin{aligned}
 d_N &= Q_N(3) - Q_N(1) \\
 &= 0,6745 - (-0,6745) \\
 &= 1,349, \\
 d_G &= Q_G(3) - Q_G(1) \\
 &= 0,39424 - (-0,39424) \\
 &= 0,78848,
 \end{aligned}$$

où $Q_N(3)$ et $Q_N(1)$ sont les troisième et premier quartiles de la loi normale centrée réduite et $Q_G(3)$ et $Q_G(1)$ sont les troisième et premier quartiles de la loi GEP de type V avec paramètres $\gamma = \delta = 0$, $\alpha = 1$, $\lambda = 4$ et $z_0 = 2$. Alors τ_0 est tel que

$$\begin{aligned}
 d_G \tau_0 &= d_N \\
 \Leftrightarrow \tau_0 &= \frac{d_N}{d_G} \\
 \Leftrightarrow \tau_0 &= 1,7109.
 \end{aligned}$$

On voit sur la figure 2.9 qu'avec ce choix, la distribution de la variable aléatoire $\frac{\varepsilon_i}{2,5}$ a une dispersion semblable à celle d'une normale centrée réduite, sauf pour les ailes. Puisque nous considérons plutôt la variable aléatoire $Z_i = \frac{Y_i}{X_i}$, alors par l'équation (2.2), la densité de cette variable évaluée au point $z_i = \frac{y_i}{x_i}$, est donnée par

$$\frac{1}{1/x_i} f_\varepsilon \left(\frac{z_i - \beta}{1/x_i} \right) \propto \left(\left| \frac{z_i - \beta}{4,2773/x_i} \right| + 2 \right)^{-1} \left(\log \left(\left| \frac{z_i - \beta}{4,2773/x_i} \right| + 2 \right) \right)^{-4}, \quad (2.10)$$

pour $i = 1, \dots, 10$ avec $z_i \in \mathbb{R}$ et $2,5 \tau_0 = 4,2773$. Nous noterons la loi de $Z_i|X_i, \beta$ par F_{GEP} .

Un problème survient avec l'estimation de β si on utilise l'espérance *a posteriori* de β , puisqu'elle n'existe pas. Cela provient du fait que les moments d'une variable aléatoire distribuée selon une GEP de type V n'existent pas (Desgagné et Angers, 2005). Ainsi, l'estimateur $b = \int_{-\infty}^{\infty} \beta \pi(\beta|Z_1, \dots, Z_{10}) d\beta$ n'existe pas. L'étude de la densité *a posteriori* $\pi(\beta|\underline{z}, \underline{z})$ à partir des observations du tableau 2.1 nous a permis de constater qu'on

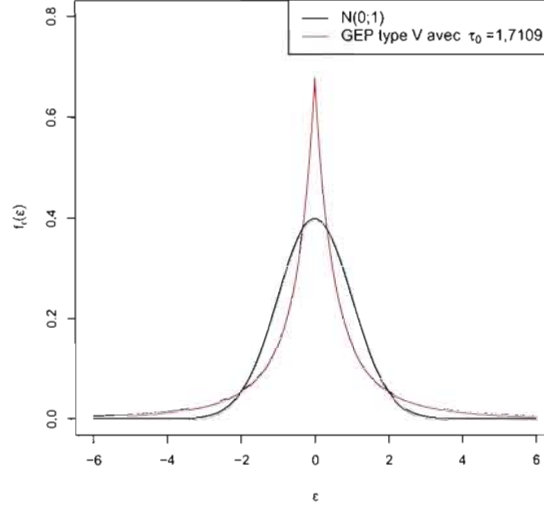


Figure 2.9 Densités $N(0;1)$ et F , où F est une loi GEP de type V avec paramètres $\gamma = \delta = 0$, $\alpha = 1$, $\lambda = 4$, $z_0 = 2$ et paramètre d'échelle $\tau_0 = 1,7109$.

peut trouver un domaine d'intégration assez grand et fini, $\mathfrak{D} = \{\beta | d_{inf} \leq \beta \leq d_{sup}\}$, où d_{inf} et d_{sup} sont les bornes d'intégration inférieure et supérieure, telles que la presque totalité de la masse de la densité s'y trouve. Par exemple, pour $d_{inf} = 0$ et $d_{sup} = 5$, la masse comprise entre $[-5000; 0)$ et $(5; 5000]$ est négligeable et 99,998% de la masse se trouve entre 0 et 5. Ainsi, le paramètre β sera estimé par

$$b = \int_{\mathfrak{D}} \beta \pi(\beta | \underline{x}, \underline{z}) d\beta.$$

Notons que cela correspond à choisir une densité *a priori* avec un domaine \mathfrak{D} , par exemple $\pi(\beta) \equiv \mathbb{I}[\beta \in \mathfrak{D}]$, où $\mathbb{I}(\cdot)$ est la fonction indicatrice, et considérer ensuite l'espérance *a posteriori* puisqu'elle existe maintenant. Notons également que le choix de \mathfrak{D} est très robuste à partir d'un certain seuil, au point où on a l'impression que l'espérance converge lors des calculs numériques. Cette dernière condition de stabilité est nécessaire pour utiliser l'espérance «tronquée» sur \mathfrak{D} .

Considérons d'abord le cas où le point P_{10} est aberrant en ordonnée. Les points $P_{10}^{(I)} =$

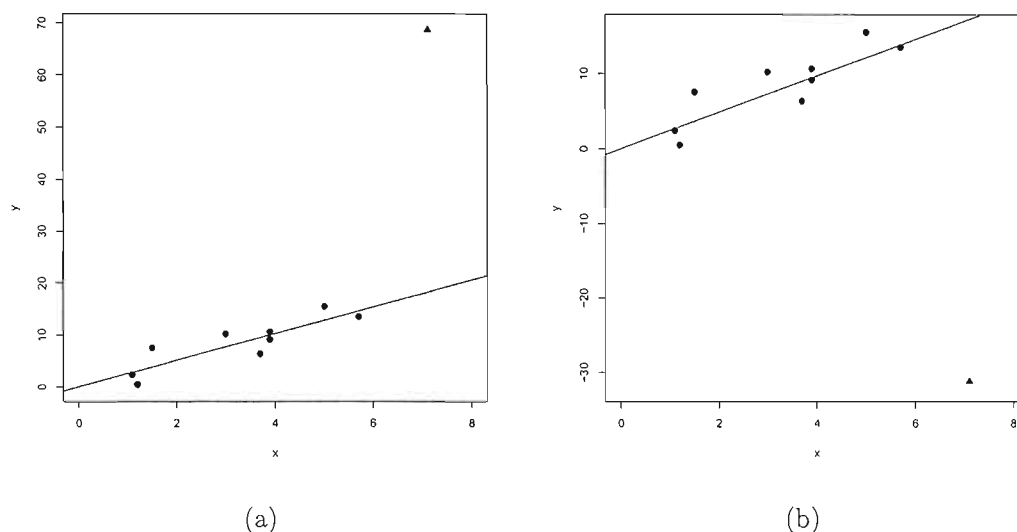


Figure 2.10 Estimé bayésien de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(I)} = (7,1; 68,8)$, (b) lorsque le point $P_{10}^{(II)} = (7,1; -31,2)$. Le symbole \blacktriangle représente le point P_{10} dont la coordonnée y a été modifiée. Les variables aléatoires Z_i suivent une loi F_{GEP} .

$(7,1; 68,8)$ et $P_{10}^{(II)} = (7,1; -31,2)$ tout comme les droites dont la pente est estimée à $b_I = 2,5704$ et à $b_{II} = 2,5257$ respectivement, sont représentés sur les graphiques (a) et (b) de la figure 2.10.

Lorsque y_{10} tend vers plus ou moins l'infini, l'estimé du paramètre β tend vers la valeur 2,5478 comme on peut le constater sur la figure 2.11. D'ailleurs, cette valeur correspond exactement à l'estimé de β lorsque le point P_{10} est retiré de l'ensemble de données, noté b_{GEP} . La méthode est donc robuste à une valeur aberrante en y .

Afin de mieux comprendre, considérons les graphiques (a), (b) et (c) de la figure 2.12, représentant les densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ pour trois valeurs positives de y_{10} . Pour la figure (a), on a considéré le point original $P_{10} = (7,1; 18,8)$, soit $z_{10} = 2,6479$. Puisque le point P_{10} n'est pas aberrant, il n'y a pas

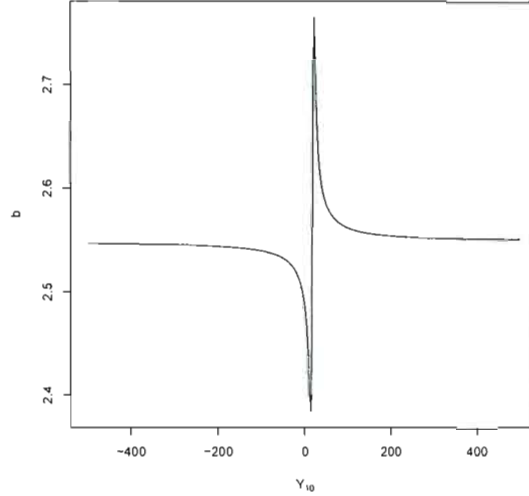


Figure 2.11 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} .

de conflit entre les points P_1, \dots, P_9 et P_{10} . Ce dernier se retrouve dans la trajectoire linéaire formée par les neuf autres points et n'est pas écarté dans le calcul de l'estimation de β . Ce cas est très similaire à celui observé à la section 2.6.1, c'est-à-dire lorsque Z_i est distribuée selon une $t_4(\beta; \frac{\sigma^2 \tau_0}{x_i})$. On voit dans la figure (a) que l'absence de conflit entre les données se traduit par la superposition des densités $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$. La densité $\pi(\beta|Z_1, \dots, Z_{10})$ est alors calculée en fonction du meilleur compromis entre les sources d'information, ce qui résulte par la superposition des trois densités.

Nous considérons maintenant la figure 2.12 (b). Le point P_{10} considéré a comme coordonnées $(7,1; 28,8)$, soit $z_{10} = 4,0563$. Il n'y a pas d'information conflictuelle, mais tout comme dans le cas observé à la section 2.6.1, on voit que le changement dans la coordonnée y_{10} a entraîné un déplacement de la densité $\pi(\beta|Z_{10})$ vers la droite de telle sorte que les densités $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ se chevauchent maintenant dans leur aile de sorte que leur zone commune est beaucoup plus petite qu'elle ne l'était à la figure 2.12 (a). La densité *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$ calculée en conséquence, voit

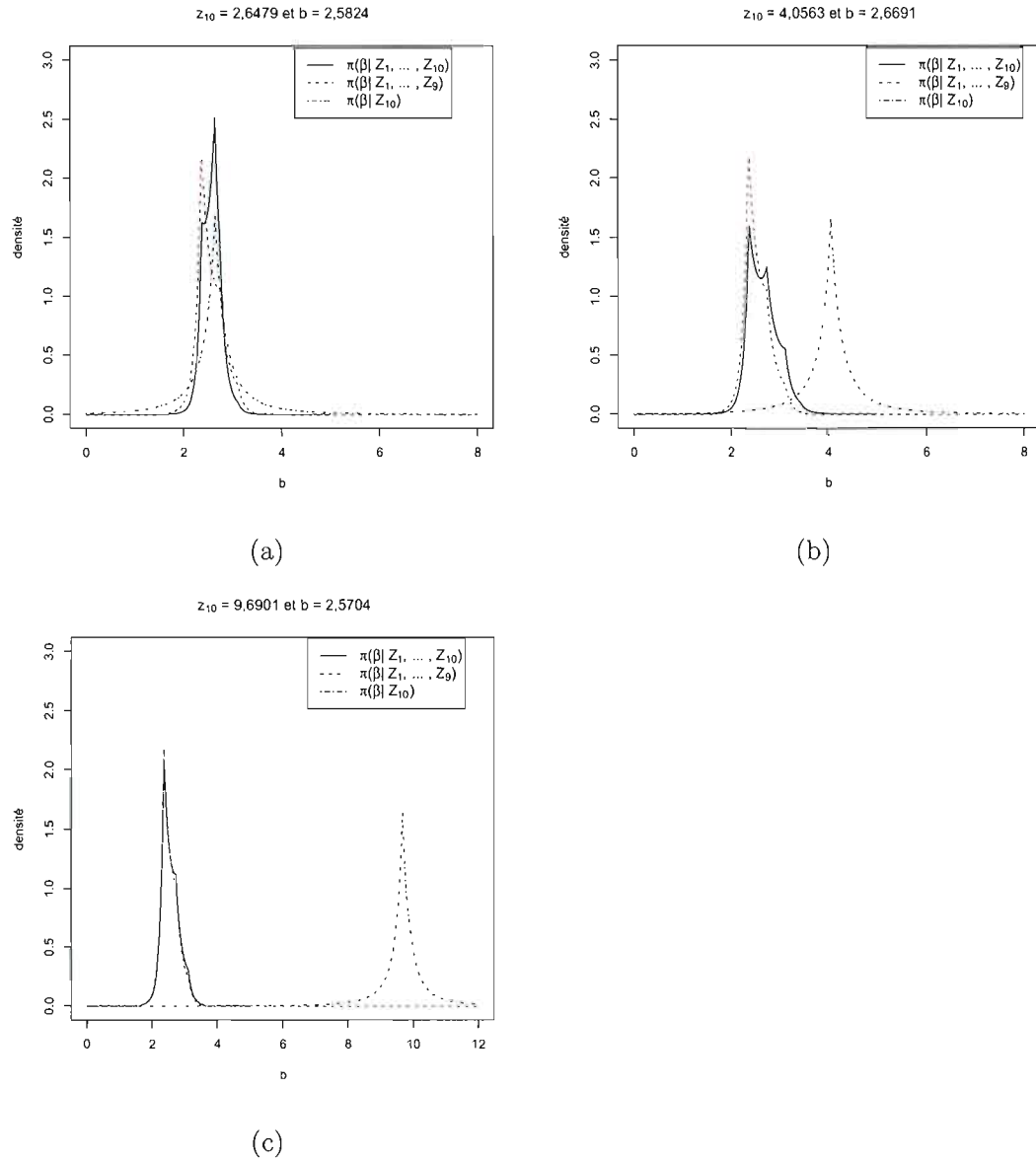


Figure 2.12 Comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque (a) $y_{10} = 18,8$, (b) $y_{10} = 28,8$ et (c) $y_{10} = 68,8$.

sa dispersion augmenter vers la droite, c'est-à-dire vers la densité *a posteriori* du point P_{10} . Cela affecte le calcul de l'espérance *a posteriori* «tronquée» résultant en un estimé de β égal à 2,669, soit une surestimation de $b_{GEP} = 2,5478$.

La figure 2.12 (c) correspond au cas où le point $P_{10} = (7,1; 68,8)$, ce qui donne $z_{10} = 9,6901$. Ce point est considéré aberrant par rapport aux points P_1, \dots, P_9 . Il s'agit d'une source d'information conflictuelle. La densité $\pi(\beta|Z_{10})$ s'est éloignée vers la droite par rapport à sa position sur la figure 2.12 (b). Du même coup, la zone commune entre les densités $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ est maintenant négligeable. Les densités $\pi(\beta|Z_1, \dots, Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ sont pratiquement superposées, ce qui résulte en un estimé de β égal à 2,5704, soit une valeur légèrement supérieure à $b_{GEP} = 2,5478$. Ces résultats concordent avec la théorie de la section 2.5, indiquant que le point P_{10} voit son influence disparaître à mesure que y_{10} tend vers plus ou moins l'infini.

Considérons maintenant le cas où le point P_{10} est aberrant en abscisse. Malgré le fait que les résultats sur la robustesse discutés à la section 2.5 ne s'appliquent pas, en raison de la variabilité de la variance de $Z_{10}|X_i, \beta$ lorsque le point P_{10} tend vers plus ou moins l'infini, nous explorons tout de même la robustesse du paramètre β . Les graphiques (a) et (b) de la figure 2.13 illustrent l'allure du nuage de points lorsque le point aberrant a comme coordonnées $P_{10}^{(III)} = (57,1; 18,8)$ et $P_{10}^{(IV)} = (-42,9; 18,8)$, respectivement. Les estimés du paramètre β sont donnés par $b_{III} = 2,4818$ et $b_{IV} = 2,4999$. Il s'agit de valeurs sous-estimant la valeur de $b_{GEP} = 2,5478$ lorsque le point P_{10} est retiré de l'ensemble de données. La figure 2.14 illustre que l'estimé de β , et donc que le point P_{10} , est de plus en plus écarté dans l'estimation de β à mesure que x_{10} tend vers plus ou moins l'infini.

À nouveau, nous allons étudier le comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ pour quatre points aberrants dont l'abscisse x_{10} égal à 57,1 ($z_{10} = 0,3292$), 1007,1 ($z_{10} = 0,0187$), 2007,1 ($z_{10} = 0,0094$) et 10007,1 ($z_{10} = 0,0019$) en considérant les figures 2.15 (a), (b), (c) et (d) respectivement.

Dans la figure 2.15 (a), le point aberrant correspond au point $P_{10}^{(III)} = (57,1; 18,8)$. On

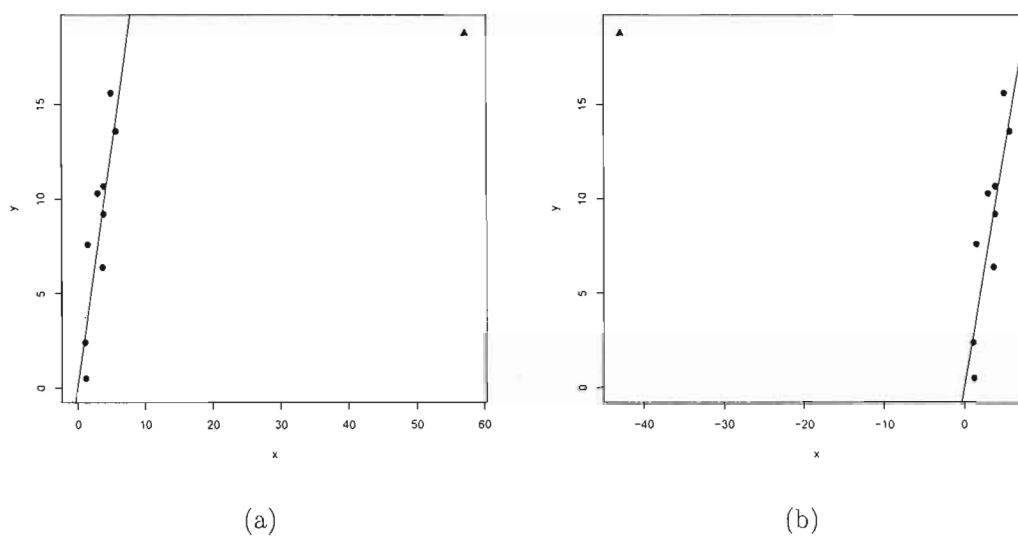


Figure 2.13 Estimé bayésien de la pente du modèle $y_i = \beta x_i + \varepsilon_i$ (a) lorsque le point $P_{10}^{(III)} = (57,1; 18,8)$, (b) lorsque le point $P_{10}^{(IV)} = (-42,9; 18,8)$. Le symbole ▲ représente le point P_{10} dont la coordonnée x a été modifiée. Les variables aléatoires Z_i suivent une loi F_{GEP} .

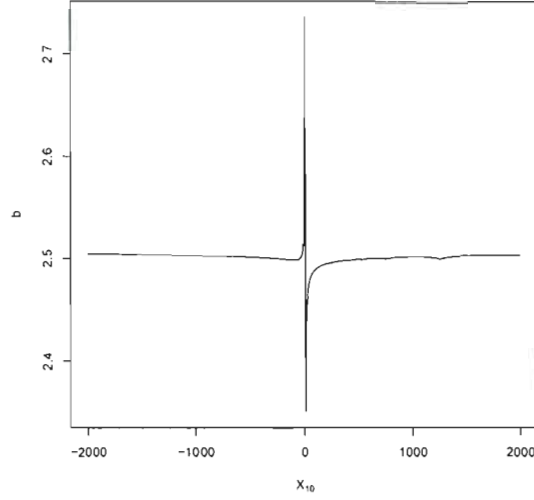


Figure 2.14 Comportement de l'estimé b pour différentes valeurs de la coordonnée y_{10} .

à vu que l'estimé $b_{III} = 2,4818$. On voit que les densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ sont pratiquement superposées. La zone commune entre les densités $\pi(\beta|Z_1, \dots, Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ semble négligeable. Par contre, puisque $b_{GEP} = 2,5478$, on constate que l'estimé b_{III} le sous-estime dû au fait que la méthode n'écarte pas totalement le point $P_{10}^{(III)}$ qui détient encore une influence dans le calcul de l'estimé de β .

Dans la figure 2.15 (b), le point aberrant considéré a comme coordonnées (1007,1 ; 18,8). La valeur aberrante et les neuf points P_1, \dots, P_9 sont conflictuels. Tel qu'on pourrait s'y attendre d'une méthode robuste, les densités $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ sont pratiquement superposées. Cependant, on observe que la densité *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$ possède un mode se trouvant dans la région de la densité $\pi(\beta|Z_{10})$. Cela pourrait laisser présager que le point aberrant P_{10} gagne en importance, mais il n'en est rien. L'estimé de β , égal à 2,5013, s'est rapproché de son point de convergence lorsque P_{10} est retiré des données, démontrant que le point aberrant possède moins d'influence que dans le

cas illustré à la figure 2.15 (a).

Maintenant, le point aberrant considéré dans le cas représenté à la figure 2.15 (c) a comme coordonnées (2007,1 ; 18,8). L'estimé de β correspondant à ce cas est 2,5026. À nouveau, la quasi-superposition des densités $\pi(\beta|Z_1, \dots, Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$, lorsqu'il y a présence d'information conflictuelle, est en accord avec la théorie de la section 2.5. Par contre, le mode se trouvant dans la région de la densité $\pi(\beta|Z_{10})$, détecté à la figure 2.15 (b), est toujours présent. Cependant, on constate que l'estimé de β a continué de converger vers $b_{GEP} = 2,5478$, démontrant que sans complètement écarter le point aberrant, la méthode lui accorde de moins en moins d'importance dans l'estimation du paramètre β .

Enfin, le point aberrant considéré dans le cas représenté à la figure 2.15 (d) a comme coordonnées (10007,1 ; 18,8). La valeur de l'estimé de β calculée correspond à 2,5039, soit une légère augmentation par rapport aux estimés obtenus dans les contextes des figures 2.15 (a), (b) et (c). À nouveau, les densités $\pi(\beta|Z_1, \dots, Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_9)$ sont presque totalement superposées, comme le prévoit la théorie lorsqu'il y a présence d'une valeur aberrante dans l'ensemble de données. Le mode de la densité *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$, se trouvant dans la région de la densité $\pi(\beta|Z_{10})$, a augmenté à un point tel qu'il est maintenant le mode global de la densité. À nouveau, cela suggère que l'influence du point aberrant P_{10} augmente. Si dans le cas classique, ce résultat apparaît comme étant vrai, car le mode *a posteriori* correspond à l'estimateur de vraisemblance maximal, on voit que l'espérance *a posteriori* comme estimateur bayésien de β est au contraire, de moins en moins influencé par le point aberrant. La méthode bayésienne, lorsque la variable $Z_i|X_i, \beta$ est distribuée selon une loi F_{GEP} produit donc un estimateur robuste à une valeur aberrante tant en ordonnée qu'en abscisse. Notons que dans cette section, afin d'utiliser une GEP de type V, nous avons dû contraindre ses paramètres γ, δ et α à prendre les valeurs 0,0 et 1 respectivement. Pour le paramètre $\lambda > 1$, nous avons choisi de façon arbitraire, $\lambda = 4$. Or, ce paramètre intervient dans la robustesse de l'estimateur de β . Ainsi, il est probable que d'autres choix de λ ait un impact plus ou moins négligeable sur la robustesse de l'estimateur de β .

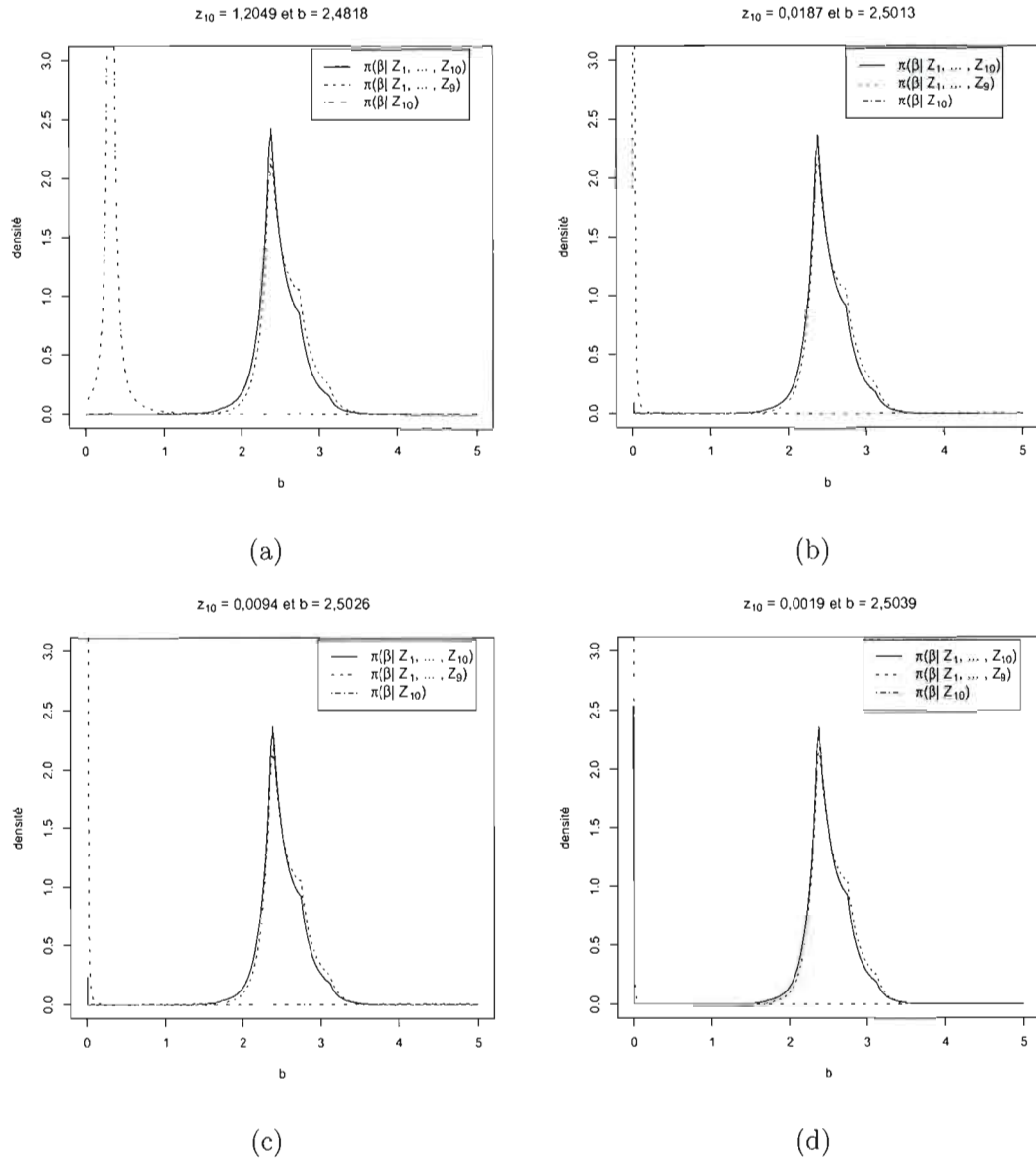


Figure 2.15 Comportement des densités *a posteriori* $\pi(\beta|Z_1, \dots, Z_9)$, $\pi(\beta|Z_{10})$ et $\pi(\beta|Z_1, \dots, Z_{10})$ lorsque (a) $x_{10} = 57,1$, (b) $x_{10} = 1007,1$, (c) $x_{10} = 2007,1$ et (d) $x_{10} = 10007,1$.

CHAPITRE III

MÉTHODE D'ESTIMATION BAYÉSIENNE AVEC PARAMÈTRE D'ÉCHELLE ALÉATOIRE

3.1 Contexte théorique

Dans les deux premiers chapitres, nous avons vu différentes méthodes permettant d'estimer le paramètre β de l'équation

$$y_i = \beta x_i + \varepsilon_i, \quad (3.1)$$

basées soit sur un contexte de statistique classique (chapitre 1), soit sur un contexte bayésien (chapitre 2). De plus, nous avons vu que certaines méthodes fréquentistes du chapitre 1 permettent aussi l'estimation d'un paramètre d'échelle σ . Au chapitre 2, la méthode d'estimation bayésienne discutée était orientée sur l'estimation du paramètre β en considérant le paramètre d'échelle connu. La robustesse de l'estimateur bayésien du paramètre d'échelle lorsque le paramètre de position est connu, en adaptant la théorie vue au chapitre 2, grâce entre autres à une variante de la p-crédence, a aussi été étudié par Desgagné (2005). Il semble alors naturel de vouloir adapter ces deux méthodes d'estimation afin d'être en mesure d'estimer simultanément les paramètres de position et d'échelle. Cependant, la théorie entourant une telle méthode d'estimation est présentement en cours de développement. Par conséquent, dans ce chapitre, la théorie entourant la robustesse des estimateurs des paramètres β et σ ne sera pas abordée. Ce chapitre sera plutôt axé sur l'exploration des résultats de robustesse de façon numérique en observant le comportement des estimés de β et σ , lorsqu'il y a présence d'une valeur

aberrante dans l'échantillon de données.

3.2 Contexte bayésien

On considère un échantillon de variables aléatoires X_1, \dots, X_n . Comparativement à la section 2.2 du chapitre 2 où nous avons considéré un seul paramètre, dans cette section, nous supposons $k \leq n$ paramètres inconnus et indépendants entre eux $\underline{\theta} = (\theta_1, \dots, \theta_k)$. Nous supposons que les variables $X_i|\underline{\theta}$, pour $i = 1, \dots, n$, sont conditionnellement indépendantes entre elles. La densité conditionnelle de chaque variable $X_i|\underline{\theta}$ est donnée par $f_i(x_i|\underline{\theta})$, où $\theta_j \in \Theta$ pour $j = 1, \dots, k$ et $X_i \in \Omega$, deux sous-ensembles de \mathbb{R} . L'approche bayésienne ne diffère pas de celle discutée à la section 2.2 du chapitre 2 en ce sens que nous mettons à jour de l'information *a priori* par de l'information contenue dans les données grâce à l'application du théorème de Bayes. Par contre, il est maintenant possible d'avoir autant de sources d'information *a priori* qu'il y a de paramètres à estimer. La densité conjointe *a posteriori* de $\underline{\theta}$ est donnée par

$$\pi(\underline{\theta}|\underline{x}) = \frac{\pi(\theta_1) \dots \pi(\theta_k) \prod_{i=1}^n f_i(x_i|\underline{\theta})}{m(\underline{x})},$$

où $\pi(\theta_1), \dots, \pi(\theta_k)$ sont les k densités *a priori* associées aux paramètres $\theta_1, \dots, \theta_k$, lesquels sont supposés être indépendants entre eux *a priori*, $\underline{x} = (x_1, x_2, \dots, x_n)$ et $m(\underline{x}) = \int_{\theta_1 \in \Theta} \dots \int_{\theta_k \in \Theta} \pi(\theta_1) \dots \pi(\theta_k) \prod_{i=1}^n f_i(x_i|\underline{\theta}) d\theta_1 \dots d\theta_k$ est la densité marginale de \underline{X} .

3.3 Le problème de position

La méthode d'estimation du paramètre β de l'équation $y_i = \beta x_i + \varepsilon_i$ et du paramètre σ , dans le contexte bayésien, requiert le calcul de la densité *a posteriori* conjointe de β et σ . Nous allons calculer cette densité de façon à exprimer le problème d'estimation de β et σ en un problème d'estimation des paramètres de position et échelle, c'est-à-dire β et σ , respectivement. Soit un ensemble de données $(x_i; y_i)$ pour $i = 1, \dots, n$, pour lequel nous supposons que la relation liant les variables X_i et Y_i est donnée par l'équation (3.1). Nous supposons que les variables X_i proviennent d'une distribution

quelconque qui ne dépend pas de β , ni de σ . Le contexte bayésien dans lequel nous travaillons nous permet de considérer Y_i , ε_i de même que les paramètres β et σ comme des variables aléatoires. Nous supposons aussi que les variables ε_i sont indépendamment et identiquement distribuées et que leur distribution est généralement choisie telle qu'elle soit symétrique sur \mathbb{R} et centrée en zéro. La densité de ε_i évaluée au point ε_i est donnée par $\frac{1}{\sigma} f_\varepsilon\left(\frac{\varepsilon_i}{\sigma}\right)$ où σ est un paramètre d'échelle inconnu. La densité conjointe *a posteriori* de β et σ est donnée par

$$\begin{aligned}\pi(\beta, \sigma | \underline{x}, \underline{y}) &= \frac{\pi(\beta) \pi(\sigma) f_{X, Y}(\underline{x}, \underline{y} | \beta, \sigma)}{m(\underline{x}, \underline{y})} \\ &\propto \pi(\beta) \pi(\sigma) f_Y(\underline{y} | \underline{x}, \beta, \sigma) f_X(\underline{x} | \beta, \sigma) \\ &\propto \pi(\beta) \pi(\sigma) f_Y(\underline{y} | \underline{x}, \beta, \sigma) f_X(\underline{x}) \\ &\propto \pi(\beta) \pi(\sigma) f_Y(\underline{y} | \underline{x}, \beta, \sigma).\end{aligned}$$

Puisque $Y_i = \beta X_i + \varepsilon_i$, et que ε_i sont indépendantes entre elles, on voit bien que $Y_i | X_i, \beta, \sigma$ sont aussi indépendantes pour $i = 1 \dots, n$. On a donc

$$\pi(\beta, \sigma | \underline{x}, \underline{y}) \propto \pi(\beta) \pi(\sigma) \prod_{i=1}^n f_{Y_i}(y_i | x_i, \beta, \sigma).$$

Par des techniques de changements de variables, puisque $\varepsilon_i = Y_i - \beta x_i$, chaque variable Y_i a comme densité conditionnelle, sachant X_i, β et σ

$$\begin{aligned}f_{Y_i}(y_i | x_i, \beta, \sigma) &= \frac{1}{\sigma} f_\varepsilon\left(\frac{y_i - \beta x_i}{\sigma}\right) \left| \frac{d\varepsilon_i}{dY_i} \right| \\ &= \frac{1}{\sigma} f_\varepsilon\left(\frac{y_i - \beta x_i}{\sigma}\right).\end{aligned}$$

On a alors

$$\begin{aligned}\pi(\beta, \sigma | \underline{x}, \underline{y}) &\propto \pi(\beta) \pi(\sigma) \prod_{i=1}^n f_{Y_i}(y_i | x_i, \beta, \sigma) \\ &\propto \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_\varepsilon\left(\frac{y_i - \beta x_i}{\sigma}\right).\end{aligned}$$

Si on définit la variable aléatoire $Z_i = Y_i/X_i$, ce qui est équivalent à $\varepsilon_i = \frac{Z_i - \beta}{1/X_i}$, nous pourrons alors considérer β comme un paramètre de position pour les variables aléatoires Z_i , considérant X_1, \dots, X_n connues. En effet, la densité de Z_i , sachant X_i, β et σ est

donnée par

$$\begin{aligned} g_{Z_i}(z_i|x_i, \beta, \sigma) &= \frac{1}{\sigma} f_\varepsilon \left(\frac{z_i - \beta}{\sigma/x_i} \right) \left| \frac{d\varepsilon_i}{dZ_i} \right| \\ &= \frac{1}{\sigma/x_i} f_\varepsilon \left(\frac{z_i - \beta}{\sigma/x_i} \right). \end{aligned} \quad (3.2)$$

On remarque que les distributions de $Y_i|X_i, \beta, \sigma$ ne diffèrent que par un paramètre de position (βx_i), tandis que les distributions de $Z_i|X_i, \beta, \sigma$ ne diffèrent que par un paramètre d'échelle (σ/x_i), tout en partageant le même paramètre de position β . Si on considère les observations $z_1 = y_1/x_1, \dots, z_n = y_n/x_n$, on a alors

$$\begin{aligned} \pi(\beta, \sigma | \underline{x}, \underline{z}) &\propto \pi(\beta) \pi(\sigma) \prod_{i=1}^n g_{Z_i}(z_i|x_i, \beta, \sigma) \\ &\propto \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma/x_i} f_\varepsilon \left(\frac{z_i - \beta}{\sigma/x_i} \right), \end{aligned}$$

puisque $Z_i|X_i, \beta, \sigma$ sont conditionnellement indépendantes, du fait que $Y_i|X_i, \beta, \sigma$ le sont aussi et que $Z_i = Y_i/X_i$, pour $i = 1, \dots, n$.

Si nous considérons X_1, \dots, X_n implicitement connues et si nous posons

$$\frac{1}{\sigma} f_i \left(\frac{z_i - \beta}{\sigma} \right) = g_{Z_i}(z_i|x_i, \beta, \sigma) = \frac{1}{\sigma/x_i} f_\varepsilon \left(\frac{z_i - \beta}{\sigma/x_i} \right),$$

où les densités f_i ne diffèrent que par un paramètre d'échelle $1/x_i$ connu, alors on est bien en présence d'un problème d'estimation des paramètres de position et échelle avec $Z_i|X_i, \beta, \sigma$ conditionnellement indépendantes ayant comme densité $\frac{1}{\sigma} f_i \left(\frac{z_i - \beta}{\sigma} \right)$. Ainsi, la densité conjointe *a posteriori* de β et σ est donnée par

$$\pi(\beta, \sigma | \underline{z}) \propto \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_i \left(\frac{z_i - \beta}{\sigma} \right).$$

3.4 Exemples

Nous considérons le jeu de données simulé à partir du modèle $y_i = 2,5 x_i + \varepsilon_i$, avec $\varepsilon_i \sim N(0; 2,5^2)$ introduit à la section (1.2). Ces données sont présentées à nouveau dans le tableau 3.1.

Dans ce chapitre, nous considérons les paramètres β et σ aléatoires. Nous les estimons par l'espérance *a posteriori* de β et σ , respectivement. Si la densité conjointe *a posteriori*

i	x_i	y_i
1	1,1	2,4
2	1,2	0,5
3	1,5	7,6
4	3,0	10,3
5	3,7	6,4
6	3,9	9,2
7	3,9	10,7
8	5,0	15,6
9	5,7	13,6
10	7,1	18,8

Tableau 3.1 Coordonnées x et y pour les 10 points de l'ensemble de données

de β et σ déterminée à la section 3.3 est donnée par

$$\pi(\beta, \sigma | \underline{z}) = \frac{\pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_i \left(\frac{z_i - \beta}{\sigma} \right)}{m(\underline{z})},$$

alors on définit les estimateurs de β et σ pour $\beta \in \mathbb{R}$ et $0 < \sigma < \infty$, par

$$\begin{aligned} b = \mathbb{E}[\beta | \underline{z}] &= \int_{-\infty}^{\infty} \int_0^{\infty} \beta \pi(\beta, \sigma | \underline{z}) d\sigma d\beta \\ &= \frac{\int_{-\infty}^{\infty} \int_0^{\infty} \beta \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_i \left(\frac{z_i - \beta}{\sigma} \right) d\sigma d\beta}{\int_{-\infty}^{\infty} \int_0^{\infty} \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_i \left(\frac{z_i - \beta}{\sigma} \right) d\sigma d\beta}, \end{aligned}$$

$$\begin{aligned} \hat{\sigma} = \mathbb{E}[\sigma | \underline{z}] &= \int_0^{\infty} \int_{-\infty}^{\infty} \sigma \pi(\beta, \sigma | \underline{z}) d\beta d\sigma \\ &= \frac{\int_0^{\infty} \int_{-\infty}^{\infty} \sigma \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_i \left(\frac{z_i - \beta}{\sigma} \right) d\beta d\sigma}{\int_0^{\infty} \int_{-\infty}^{\infty} \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_i \left(\frac{z_i - \beta}{\sigma} \right) d\beta d\sigma}, \end{aligned}$$

respectivement. Si on considère la notation de la régression linéaire, alors

$$\begin{aligned} b = \mathbb{E}[\beta | \underline{z}, \underline{x}] &= \int_{-\infty}^{\infty} \int_0^{\infty} \beta \pi(\beta, \sigma | \underline{z}, \underline{x}) d\sigma d\beta \\ &= \frac{\int_{-\infty}^{\infty} \int_0^{\infty} \beta \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma/x_i} f_{\varepsilon} \left(\frac{z_i - \beta}{\sigma/x_i} \right) d\sigma d\beta}{\int_{-\infty}^{\infty} \int_0^{\infty} \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma/x_i} f_{\varepsilon} \left(\frac{z_i - \beta}{\sigma/x_i} \right) d\sigma d\beta}, \end{aligned}$$

$$\begin{aligned} \hat{\sigma} = \mathbb{E}[\sigma | \underline{z}, \underline{x}] &= \int_0^{\infty} \int_{-\infty}^{\infty} \sigma \pi(\beta, \sigma | \underline{z}, \underline{x}) d\beta d\sigma \\ &= \frac{\int_0^{\infty} \int_{-\infty}^{\infty} \sigma \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma/x_i} f_i \left(\frac{z_i - \beta}{\sigma} \right) d\beta d\sigma}{\int_0^{\infty} \int_{-\infty}^{\infty} \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma/x_i} f_i \left(\frac{z_i - \beta}{\sigma} \right) d\beta d\sigma}, \end{aligned}$$

ou de façon équivalente,

$$\begin{aligned} b = \mathbb{E}[\beta | \underline{x}, \underline{y}] &= \int_{-\infty}^{\infty} \int_0^{\infty} \beta \pi(\beta, \sigma | \underline{x}, \underline{y}) d\sigma d\beta \\ &= \frac{\int_{-\infty}^{\infty} \int_0^{\infty} \beta \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_{\varepsilon} \left(\frac{y_i - \beta x_i}{\sigma} \right) d\sigma d\beta}{\int_{-\infty}^{\infty} \int_0^{\infty} \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_{\varepsilon} \left(\frac{y_i - \beta x_i}{\sigma} \right) d\sigma d\beta}, \end{aligned}$$

$$\begin{aligned} \hat{\sigma} = \mathbb{E}[\sigma | \underline{x}, \underline{y}] &= \int_0^{\infty} \int_{-\infty}^{\infty} \sigma \pi(\beta, \sigma | \underline{x}, \underline{y}) d\beta d\sigma \\ &= \frac{\int_0^{\infty} \int_{-\infty}^{\infty} \sigma \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_{\varepsilon} \left(\frac{y_i - \beta x_i}{\sigma} \right) d\beta d\sigma}{\int_0^{\infty} \int_{-\infty}^{\infty} \pi(\beta) \pi(\sigma) \prod_{i=1}^n \frac{1}{\sigma} f_{\varepsilon} \left(\frac{y_i - \beta x_i}{\sigma} \right) d\beta d\sigma}. \end{aligned}$$

Afin d'essayer d'être le plus neutre possible quant au choix des densités *a posteriori* de β et σ , nous avons choisi $\pi(\beta) \equiv 1$ et $\pi(\sigma) = \frac{1}{\sigma}$. Il s'agit de deux distributions impropres et non informatives, typiquement employées pour des problèmes de position et échelle.

Alors qu'au chapitre 2 nous avons traité les cas d'une valeur aberrante en ordonnée et en abscisse, malgré que la théorie sur la robustesse de la section 2.5 ne s'applique pas à ce dernier, dans ce chapitre nous considérons seulement le cas d'une valeur aberrante en ordonnée. Puisque la théorie concernant l'estimation simultanée des paramètres β et σ est en cours de développement, et que cette théorie suppose une variance constante pour chaque variable $Z_i = \frac{Y_i}{X_i}$, ce qui n'est pas le cas pour une valeur aberrante en abscisse, nous n'explorons pas le cas où x_{10} tend vers plus ou moins l'infini. Cependant, comme

la densité GEP de type V a des ailes plus relevées que la densité Student, tel que vu à la section 2.6 du chapitre 2, l'estimé de β en présence d'une valeur aberrante en x , lorsque le paramètre d'échelle est inconnu, devrait se comporter de façon similaire à l'estimé de β lorsque σ est considéré fixe, c'est-à-dire robuste si la variable Z_i est distribuée selon une loi F_{GEP} et non robuste si Z_i est distribuée selon une loi $t_4(\beta; \frac{2,2765}{x_i})$. Quant à l'estimé de σ lorsqu'il y a une valeur aberrante en x , il devrait se comporter de façon similaire au cas où il y a présence d'une valeur aberrante en ordonnée, c'est-à-dire robuste si la variable Z_i est distribuée selon une loi F_{GEP} et non robuste si Z_i est distribuée selon une loi $t_4(\beta; \frac{2,2765}{x_i})$.

3.4.1 Student

Dans cette section, nous supposons que la variable aléatoire ε_i est distribuée selon une loi F avec paramètre d'échelle inconnu σ , où F est une loi de Student avec 4 degrés de liberté et avec paramètre d'échelle τ_0 connu, égal à 0,9106, tel que trouvé à la section 2.6.1 du chapitre 2. Si on considère plutôt la variable aléatoire $Z_i|X_i, \beta, \sigma$, alors par l'équation (3.2), la densité de Z_i sachant X_i, β et σ est donnée par

$$\frac{1}{\sigma/x_i} f_i\left(\frac{z_i - \beta}{\sigma/x_i}\right) \propto \frac{1}{\sigma/x_i} \left(1 + \frac{1}{4} \left(\frac{z_i - \beta}{0,9106\sigma/x_i}\right)^2\right)^{-5/2},$$

pour $i = 1, \dots, 10$. Rappelons que le paramètre d'échelle connu τ_0 fait en sorte que la densité F de la variable aléatoire ε_i/σ a une dispersion semblable à celle de la loi normale centrée réduite, soit la même distance interquartile. Nous noterons la loi de $Z_i|X_i, \beta, \sigma$ par $t_4\left(\beta; \sigma\left(\frac{0,9106}{x_i}\right)\right)$, soit une Student avec quatre degrés de liberté, avec paramètre de position β et paramètre d'échelle $\sigma\left(\frac{0,9106}{x_i}\right)$.

Considérons d'abord l'estimé du paramètre β lorsque le point P_{10} est aberrant en ordonnée. Nous avons analysé la valeur de l'estimé de β pour quelques valeurs de y_{10} . Ces résultats sont présentés dans le tableau 3.2. On peut y observer que, tout comme on s'y attend d'une méthode robuste, la valeur de b tend vers une estimation réaliste de la pente d'une droite passant par les points P_1, \dots, P_9 , suggérant que la densité marginale *a posteriori* de β , $\pi(\beta|Z_1, \dots, Z_{10})$ tend vers la densité $\pi(\beta|z_1, \dots, z_9)$, comme nous

y_{10}	z_{10}	b	$\hat{\sigma}$
-3×10^8	$-4,2254 \times 10^7$	2,6172	4,6478
-2×10^8	$-2,8169 \times 10^7$	2,6172	4,6478
-200	-28,169	2,5861	4,6331
-31,2	-4,3944	2,4945	4,4944
68,8	9,6901	2,7408	4,5192
200	28,169	2,6552	4,6420
2×10^8	$2,8169 \times 10^7$	2,6172	4,6478
3×10^8	$4,2254 \times 10^7$	2,6172	4,6478
Sans P_{10}		2,6083	2,6427

Tableau 3.2 Estimés b et $\hat{\sigma}$ pour certaines valeurs de la coordonnée y_{10} ($z_{10} = y_{10}/x_{10}$), lorsque la variable $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi $t_4 \left(\beta; \sigma \left(\frac{0,9106}{x_i} \right) \right)$

l'avons vu au chapitre 2. On constate sur la figure 3.1 (a) que lorsque l'ordonnée y_{10} (z_{10}) tend vers plus ou moins l'infini, la valeur de l'estimé de β converge vers 2,6172, soit une légère surestimation de $b_S = 2,6083$, correspondant à la valeur de b lorsque le point P_{10} est retiré de l'ensemble de données. Cependant, bien que cette surestimation suggère que le point aberrant P_{10} n'est jamais complètement écarté par la méthode, l'influence qu'il détient sur la droite est limitée à un point tel que l'estimé de la pente $b = 2,6172$ est représentatif de la trajectoire linéaire formée par les points P_1, \dots, P_9 . Il semble donc tout de même approprié de qualifier la méthode d'estimation, dont la variable $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi $t_4 \left(\beta; \sigma \left(\frac{0,9106}{x_i} \right) \right)$, de méthode robuste.

Considérons maintenant l'estimé du paramètre d'échelle σ lorsque le point P_{10} est aberrant en ordonnée. Nous avons analysé la valeur de l'estimé $\hat{\sigma}$ pour certaines valeurs de y_{10} (z_{10}). Les résultats de cette analyse présentés dans le tableau 3.2 combinés à la figure 3.1 (b) représentant le comportement de l'estimé $\hat{\sigma}$, tendent à montrer que le point aberrant P_{10} a une influence limitée sur $\hat{\sigma}$, puisque ce dernier tend vers $\hat{\sigma} = 4,6478$ lorsque y_{10} (z_{10}) tend vers plus ou moins l'infini. Cependant, la méthode d'estimation,

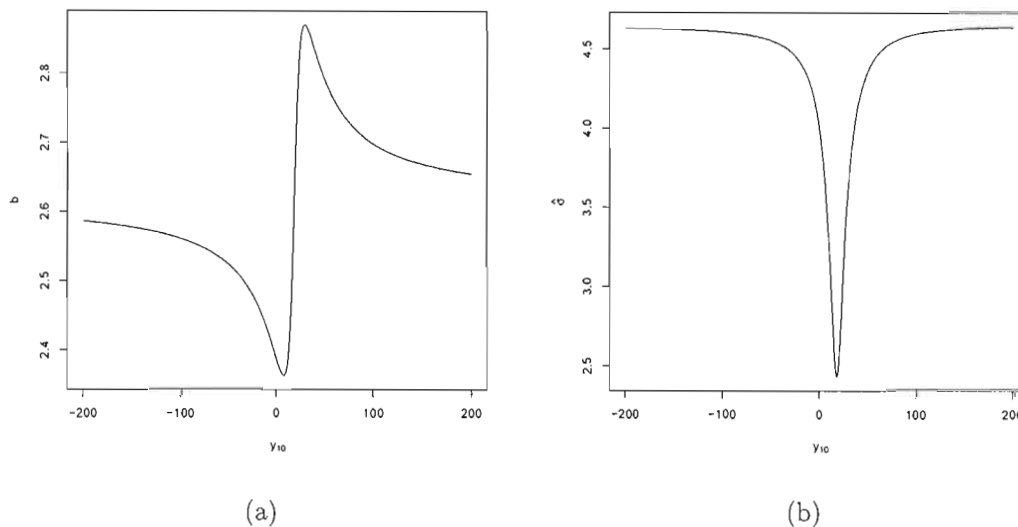


Figure 3.1 Comportement de (a) l'estimé b et de (b) l'estimé $\hat{\sigma}$ pour différentes valeurs de la coordonnée y_{10} , lorsque la variable $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi $t_4\left(\beta; \sigma\left(\frac{0.9106}{x_i}\right)\right)$.

lorsque $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi $t_4\left(\beta; \sigma\left(\frac{0.9106}{x_i}\right)\right)$, manque de finesse en surestimant la valeur de $\hat{\sigma}_S = 2,6427$, correspondant à l'estimé de σ lorsque le point P_{10} est retiré de l'ensemble de données. L'influence que détient le point aberrant P_{10} dans l'estimation de σ est plus importante que ce à quoi on s'attend d'une méthode d'estimation robuste.

La méthode d'estimation bayésienne, lorsque $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi $t_4\left(\beta; \sigma\left(\frac{0.9106}{x_i}\right)\right)$, semble robuste à une valeur aberrante en ordonnée pour un paramètre de position (et donc pour la pente β de l'équation (3.1)), mais non pour un paramètre d'échelle.

3.4.2 GEP de type V

Dans cette section, nous supposons que la variable aléatoire ε_i est distribuée selon une loi F avec paramètre d'échelle inconnu σ , où F est une loi GEP de type V avec paramètre $\gamma = \delta = 0$, $\alpha = 1$, $\lambda = 4$, $z_0 = 2$ et avec paramètre d'échelle τ_0 connu, égal à 1,7109, tel que trouvé à la section 2.6.2 du chapitre 2. Tout comme au chapitre précédent, le choix de $\lambda = 4$ est arbitraire et il est possible qu'une valeur différente pour ce paramètre puisse avoir un impact sur la robustesse de β . Si on considère plutôt la variable aléatoire $Z_i|X_i, \beta, \sigma$, alors par l'équation (3.2), la densité de Z_i sachant X_i , β et σ est donnée par

$$\frac{1}{\sigma/x_i} f_{\varepsilon_i} \left(\frac{z_i - \beta}{\sigma/x_i} \right) \propto \left(\left| \frac{z_i - \beta}{1,7109\sigma/x_i} \right| + 2 \right)^{-1} \left(\log \left(\left| \frac{z_i - \beta}{1,7109\sigma/x_i} \right| + 2 \right) \right)^{-4},$$

pour $i = 1, \dots, 10$. Rappelons que le paramètre d'échelle connu τ_0 fait en sorte que la densité F de la variable aléatoire ε_i/σ a une dispersion semblable à celle de la loi normale centrée réduite. Nous noterons la loi de $Z_i|X_i, \beta, \sigma$ par H_{GEP} .

Comme nous l'avons mentionné à la section 2.6.2 du chapitre 2, les moments de la variable $Z_i|X_i, \beta, \sigma$ n'existent pas. Tel que proposé dans cette même section, nous estimons β et σ par l'espérance *a posteriori* de β et de σ , respectivement, définie sur un domaine \mathcal{D} assez grand.

Considérons d'abord l'estimation du paramètre β lorsqu'il y a présence d'une valeur aberrante en ordonnée. Le tableau 3.3 présentant la valeur de l'estimé b pour quelques valeurs de y_{10} , ainsi que la figure 3.2 (a), indiquent que l'influence du point aberrant P_{10} sur l'estimation de β s'estompe à mesure que y_{10} (z_{10}) tend vers plus ou moins l'infini. En fait, lorsque l'ordonnée du point P_{10} tend vers plus ou moins l'infini, l'estimé du paramètre β tend vers 2,5296, soit la valeur de b lorsque le point P_{10} est soustrait de l'ensemble de données. Cela suggère qu'à la limite, la densité marginale *a posteriori* $\pi(\beta|Z_1, \dots, Z_{10})$ converge vers la densité marginale *a posteriori* $\pi(\beta|z_1, \dots, z_9)$ tel qu'indiqué par la théorie du chapitre 2. Comme on s'y attend de la part d'une méthode robuste, l'influence du point aberrant P_{10} diminue à mesure que ce dernier s'éloigne de la trajectoire linéaire formée par les points P_1, \dots, P_9 . Tout indique que la méthode

y_{10}	z_{10}	b	$\hat{\sigma}$
-1×10^{305}	$-1,4085 \times 10^{304}$	2,5297	2,4337
-1×10^{100}	$-1,4085 \times 10^{99}$	2,5299	2,4425
-3×10^8	$-4,2254 \times 10^7$	2,5330	2,5642
-2×10^8	$-2,8169 \times 10^7$	2,5331	2,5673
-200	-28,169	2,5383	3,0729
-31,2	-4,3944	2,5209	3,2103
68,8	9,6901	2,5727	3,2346
200	28,169	2,5504	3,1017
2×10^8	$2,8169 \times 10^7$	2,5331	2,5673
3×10^8	$4,2254 \times 10^7$	2,5333	2,5642
1×10^{100}	$1,4085 \times 10^{99}$	2,5299	2,4425
1×10^{305}	$1,4085 \times 10^{304}$	2,5297	2,4337
Sans P_{10}		2,5296	2,4324

Tableau 3.3 Estimés b et $\hat{\sigma}$ pour certaines valeurs de la coordonnée y_{10} ($z_{10} = y_{10}/x_{10}$), lorsque la variable $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi H_{GEP}

d'estimation bayésienne d'un paramètre de position lorsque la variable $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi H_{GEP} , est robuste à une valeur aberrante en ordonnée.

Considérons maintenant l'estimé du paramètre d'échelle σ lorsque le point P_{10} est aberrant en ordonnée. L'analyse de la figure 3.2 (b) représentant le comportement de l'estimé de σ , ainsi que du tableau 3.3 illustrant $\hat{\sigma}$ pour certaines valeurs de y_{10} semblent indiquer que le point aberrant P_{10} a une influence sur l'estimation du paramètre σ qui s'estompe à mesure que y_{10} tend vers plus ou moins l'infini. L'estimé de σ , lorsque l'ordonnée du point P_{10} tend vers plus ou moins l'infini, tend vers $\hat{\sigma}_{GEP} = 2,4324$, soit la valeur de $\hat{\sigma}$ lorsque le point P_{10} est soustrait de l'ensemble de données. Bien qu'il s'agit d'un comportement auquel on s'attend pour une méthode d'estimation robuste, la convergence de σ vers $\hat{\sigma}_{GEP}$ est très lente. Néanmoins, l'analyse numérique tend à montrer

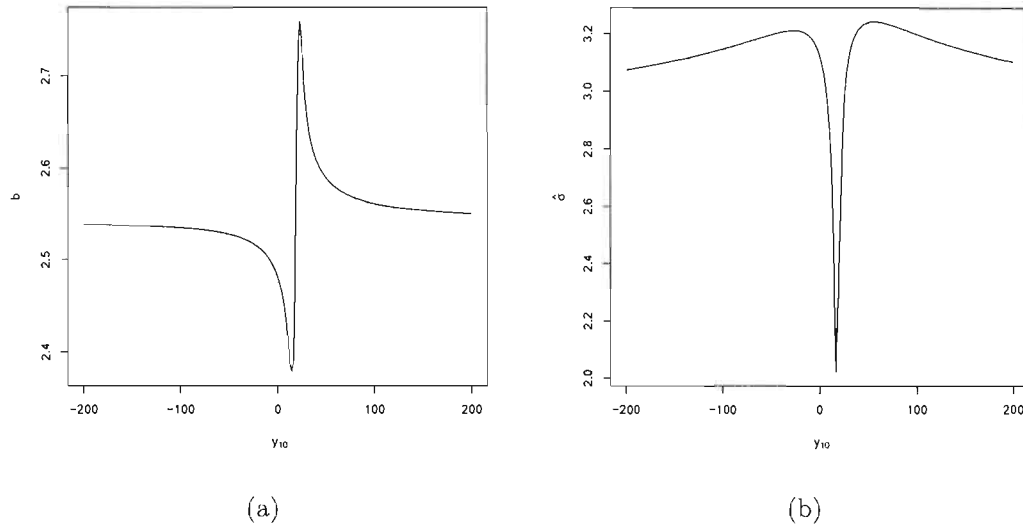


Figure 3.2 Comportement de (a) l'estimé b et de (b) l'estimé $\hat{\sigma}$ pour différentes valeurs de la coordonnée y_{10} lorsque la variable $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi H_{GEP} .

que la méthode d'estimation bayésienne d'un paramètre d'échelle lorsque $Z_i|X_i, \beta, \sigma$ est distribuée selon une loi H_{GEP} est robuste à une valeur aberrante en ordonnée.

CONCLUSION

Dans ce mémoire, nous avons étudié la robustesse d'estimateurs du paramètre β de l'équation $y_i = \beta x_i + \varepsilon_i$ et du paramètre d'échelle σ , lorsque le jeu de données contient un point aberrant. Dans un premier temps, nous nous sommes intéressés à des méthodes d'estimations basées sur l'approche fréquentiste. Sous la supposition que ε_i est i.i.d $N(0; \sigma^2)$, nous avons d'abord analysé un cas classique, c'est-à-dire l'estimateur du maximum de vraisemblance des paramètres β (correspondant à l'estimateur des moindres carrés) et σ . Nous avons vu que cette méthode d'estimation n'est pas robuste à une valeur aberrante tant en abscisse qu'en ordonnée. Nous avons alors étudié quelques alternatives à la méthode des moindres carrés afin d'estimer de façon robuste les paramètres β et σ . Parmi ces méthodes alternatives, nous avons exploré la robustesse des estimateurs-M. Les estimateurs-M se sont avérés ne pas être robuste à une valeur aberrante en abscisse et ce, tant pour le paramètre β que pour le paramètre d'échelle σ . En revanche, cette méthode d'estimation a permis de calculer des estimateurs présentant une certaine robustesse pour les paramètres β et σ , lorsqu'il y avait présence d'une valeur aberrante en ordonnée. Cependant, pour la majorité des estimateurs-M, les estimés obtenus manquaient de finesse en ce sens que, bien que l'influence de la valeur aberrante soit limitée, la méthode ne réussit pas à écarter complètement le point extrême, ce qui lui confère un certain poids dans le calcul de l'estimé. Par conséquent, la plupart des estimés obtenus, surestimaient ou sous-estimaient la valeur de l'estimé calculée lorsque le point aberrant était exclu du jeu de données. Enfin, nous avons conclu notre analyse des méthodes d'estimations d'un contexte fréquentiste par l'exploration de la technique d'estimation de la droite résistante. Contrairement aux estimateurs-M, cette méthode a produit un estimé de β robuste à une valeur aberrante tant en abscisse qu'en ordonnée. Cependant, tout comme les estimateurs-M, l'estimé obtenu manquait de finesse en surestimant ou sous-estimant la valeur de l'estimé calculée à partir de l'échantillon

de données auquel on a retiré le point aberrant.

Nous avons ensuite proposé une méthode d'estimation basée sur un contexte bayésien. Cette approche repose sur des résultats théoriques selon lesquels des densités ayant des ailes relevées sont nécessaires afin de faire de l'inférence robuste. Cependant, la théorie étant développée essentiellement pour des problèmes d'estimations d'un paramètre de position, avec paramètre d'échelle connu (et vice-versa), nous avons reformulé l'estimation de la pente de la droite $y = \beta x + \varepsilon$ en un problème d'estimation d'un paramètre de position β avec paramètre d'échelle σ connu. Afin de modéliser les ailes des densités, nous avons considéré deux lois : la loi de Student avec quatre degrés de liberté et la loi GEP de type V avec paramètres $\delta = \gamma = 0$, $\alpha = 1$, $\lambda = 4$ et $z_0 = 2$. Nous avons montré qu'une densité de Student possède des ailes plus relevées que celles d'une densité normale et qu'une densité GEP de type V possède des ailes plus relevées que celle de la densité de Student. Les ailes peu relevées de la densité normale sont à l'origine de l'estimation non robuste par la méthode des moindres carrés observée au chapitre 1. D'ailleurs, sous certaines conditions, nous avons vu que la méthode d'estimation des moindres carrés était un cas particulier de la méthode bayésienne. La méthode d'estimation bayésienne, lorsqu'on considère la densité de Student s'est avérée posséder quelques similarités avec les estimateurs-M, en ce sens que, lorsqu'il y a présence d'une valeur aberrante en ordonnée, la méthode n'écarte pas complètement le point aberrant comme on pourrait s'y attendre d'une méthode robuste, alors que la méthode ne produit pas un estimateur robuste à une valeur aberrante en abscisse. Cependant, ce léger manque de finesse observé pour la méthode bayésienne est loin d'être aussi flagrant que pour la méthode d'estimation-M. Enfin, lorsqu'on considère la densité GEP de type V, la méthode bayésienne a permis d'obtenir un estimateur du paramètre β robuste autant à une valeur aberrante en abscisse, qu'en ordonnée.

Nous avons conclu en explorant les résultats d'une estimation bayésienne simultanée des paramètres β et σ , et ce, malgré que les résultats théoriques d'une telle estimation n'aient pas encore été publiés. Les résultats d'estimation de la méthode bayésienne avec densité GEP de type V se sont avérés robustes autant pour le paramètre β que pour le

paramètre σ annonçant une approche fort prometteuse dans le domaine de la robustesse d'estimateurs de paramètres de la régression.

APPENDICE A

PROGRAMMES INFORMATIQUES

Les résultats obtenus dans ce mémoire ont été calculés avec l'aide du logiciel **R**. Il s'agit d'un progiciel disponible gratuitement sur Internet et dont le langage de programmation est très similaire à celui de S-PLUS. Ce logiciel peut être téléchargé du site internet *<http://www.r-project.org/>*. Nous allons maintenant énumérer le code de programmation nous ayant permis de réaliser les calculs pour les différents estimateurs étudiés dans ce mémoire. Dans ce qui suit, les commentaires sont précédés du caractère `#`. De plus, afin de les distinguer du code informatique, ce dernier est écrit en italique.

`#Le jeu de données.`

`y = c(2.4, 0.5, 7.6, 10.3, 6.4, 9.2, 10.7, 15.6, 13.6, 18.8)`

`x = c(1.1, 1.2, 1.5, 3, 3.7, 3.9, 3.9, 5, 5.7, 7.1)`

`#Estimateur du paramètre β par la méthode des moindres carrés avec ordonnée à l'origine $\alpha = 0$.`

`b.mc = lm(y ~ 0 + x)$coef`

`#Estimateur du paramètre d'échelle par la méthode du maximum de vraisemblance.`

```

e = y - x * b.initial
sigma.mc = sigma.initial = sqrt((1/n) * sum(e^2))

```

#Estimation des paramètres β et σ par la méthode des moindres valeurs absolues en utilisant le processus d'itération des moindres carrés repondérés. L'estimé b des moindres carrés est utilisé comme valeur initiale.

```

psi.MVA = function(e)sign(e)
b.initial = lm(y ~ 0 + x)$coef
e.k = y - x * b.initial
w.initial = (psi.MVA(e = e.k))/(e.k)
sigma.initial = mad(e.k)
n = length(x)
b.k = b.initial
w.k = w.initial
vecteur.b = c(b.initial)
vecteur.sigma = c(sigma.initial)
test = 10
compteur = 1

while(test > 0.0000001){
  b.kplus1 = (sum(x * y * w.k))/(sum((x^2) * w.k))
  vecteur.b = c(vecteur.b, b.kplus1)
  b.k = b.kplus1
  e.k = y - x * b.k
  w.k = psi.MVA(e = e.k)/(e.k)
  sigma.k = mad(e.k)
}

```

```

vecteur.sigma = c(vecteur.sigma, sigma.k)

compteur = compteur + 1

test = abs(vecteur.b[compteur]-vecteur.b[compteur-1])+abs(vecteur.sigma[compteur]-
    vecteur.sigma[compteur - 1])}

b.mva = vecteur.b[compteur]

sigma.mva = vecteur.sigma[compteur]


```

#Estimation des paramètres β et σ par la méthode de Huber en utilisant le processus d'itération des moindres carrés repondérés. L'estimé b des moindres carrés est utilisé comme valeur initiale.

```

psi.huber = fonction(e, m, sigma){ifelse(abs(e/sigma) <= m, (2 * e)/(sigma)^2,
    2 * (m/sigma) * sign(e/sigma))}

b.initial = lm(y ~ 0 + x)$coef

e.k = y - x * b.initial

sigma.initial = mad(e.k)

w.initial = (psi.huber(e = e.k, m = Constante.m, sigma = sigma.initial))/
    (e.k/sigma.initial)

n = length(x)

b.k = b.initial

sigma.k = sigma.initial

w.k = w.initial

vecteur.b = c(b.initial)

vecteur.sigma = c(sigma.initial)

,
test = 10

compteur = 1

```

```

while(test > 0.00000001){
  b.kplus1 = (sum(x * y * w.k))/(sum((x^2) * w.k))
  vecteur.b = c(vecteur.b, b.kplus1)
  b.k = b.kplus1
  e.k = y - x * b.k
  sigma.k = mad(e.k)
  w.k = psi.huber(e = e.k, m = Constante.m, sigma = sigma.k)/(e.k/sigma.k)
  vecteur.sigma = c(vecteur.sigma, sigma.k)
  compteur = compteur + 1
  test = abs(vecteur.b[compteur]-vecteur.b[compteur-1])+abs(vecteur.sigma[compteur]-
    vecteur.sigma[compteur - 1])}
  if(compteur > 5000){ #Lorsque b oscille entre deux valeurs, choisit le max
    test = 0.000000001
    b.kplus1 = max(vecteur.b[compteur], vecteur.b[compteur - 1])
    vecteur.b = c(vecteur.b, b.kplus1)}
  b.huber = vecteur.b[compteur]
  sigma.huber = vecteur.sigma[compteur]

```

#Estimation des paramètres β et σ par la méthode de Andrews en utilisant le processus d'itération des moindres carrés repondérés. L'estimé b des moindres valeurs absolues est utilisé comme valeur initiale.

```

psi.andrews = function(e, A, sigma){ifelse(abs(e/sigma) <= pi * A,
  (A/sigma) * (sin((e)/(sigma * A))), 0)}
poids.andrews = function(e, A, sigma){ifelse(e == 0, A, psi.andrews(e, A, sigma)/
  (e/sigma))}

```

```

Constante.A = 2.1

b.initial = b.mva

e.k = y - x * b.initial

sigma.initial = sigma.mva

w.initial = poids.andrews(e = e.k, A = Constante.A, sigma = sigma.initial)

n = length(x)

b.k = b.initial

sigma.k = sigma.initial

w.k = w.initial

vecteur.b = c(b.initial)

vecteur.sigma = c(sigma.initial)

test = 10

compteur = 1

while(test > 0.0000001){
  b.kplus1 = (sum(x * y * w.k)) / (sum((x^2) * w.k))
  vecteur.b = c(vecteur.b, b.kplus1)
  b.k = b.kplus1
  e.k = y - x * b.k
  sigma.k = mad(e.k)
  w.k = poids.andrews(e = e.k, A = Constante.A, sigma = sigma.k)
  vecteur.sigma = c(vecteur.sigma, sigma.k)
  compteur = compteur + 1
  test = abs(vecteur.b[compteur] - vecteur.b[compteur - 1]) + abs(vecteur.sigma[compteur] -
    vecteur.sigma[compteur - 1])}

```



```
b.andrews = vecteur.b[compteur]
```

```
sigma.andrews = vecteur.sigma[compteur]
```

#Estimation des paramètres β et σ par la méthode bipoids de Tukey en utilisant le processus d'itération des moindres carrés repondérés. L'estimé b des moindres valeurs absolues est utilisé comme valeur initiale.

```
psi.biweight = function(e, B, sigma) { ifelse(abs(e/sigma) <= B,  
  (e/sigma) * (1 - (e/(sigma * B)^2)^2), 0) }
```

```
Constante.B = 8
```

```
b.initial = b.mva
```

```
e.k = y - x * b.initial
```

```
sigma.initial = sigma.mva
```

```
w.initial = psi.biweight(e = e.k, B = Constante.B, sigma = sigma.initial) /  
  (e.k/sigma.initial)
```

```
n = length(x)
```

```
b.k = b.initial
```

```
sigma.k = sigma.initial
```

```
w.k = w.initial
```

```
vecteur.b = c(b.initial)
```

```
vecteur.sigma = c(sigma.initial)
```

```
test = 10
```

```
compteur = 1
```

```

while(test > 0.0000001){
  b.kplus1 = (sum(x * y * w.k))/(sum((x^2) * w.k))
  vecteur.b = c(vecteur.b, b.kplus1)
  b.k = b.kplus1
  e.k = y - x * b.k
  sigma.k = mad(e.k)
  w.k = psi.biweight(e = e.k, B = Constante.B, sigma = sigma.k)/(e.k/sigma.k)
  vecteur.sigma = c(vecteur.sigma, sigma.k)
  compteur = compteur + 1
  test = abs(vecteur.b[compteur]-vecteur.b[compteur-1])+abs(vecteur.sigma[compteur]-
    vecteur.sigma[compteur - 1])}

b.bipoids = vecteur.b[compteur]

sigma.bipoids = vecteur.sigma[compteur]

```

#Estimation du paramètre β par la méthode de la droite résistante de Tukey. Tout d'abord, définition de différentes fonctions. Notons que les fonctions décrites ci-dessous ont été créées pour être utilisées avec l'ensemble de données composé des vecteurs x et y . Une généralisation de ces fonctions sera nécessaire afin de pouvoir les utiliser avec tout autre échantillon de données.

#La fonction ci-dessous retourne les valeurs initiales en créant, entre autres, les sous-ensemble G et D selon l'ordre des observations du vecteur x .

```

Initial.Resis.Line <- function(x, y, n, correction, choix.resultat){
  if(x[10] < x[4]){
    y.ordon = y[order(x)] #Ordonne les observations y selon l'ordre des x.
    x.ordon = sort(x) #Ordonne les observations x en ordre croissant.

    G = cbind(x.ordon[1 : 4], y.ordon[1 : 4])
      #correspond aux valeurs originales indicées 1,2,3 et 10 (pas nécessairement dans
      cet ordre)
    C = cbind(x.ordon[5 : 6], y.ordon[5 : 6]) #correspond aux valeurs originales indicées 5
      et 6.
    D = cbind(x.ordon[7 : 10], y.ordon[7 : 10]) #correspond aux valeurs originales indicées
      6,7,8 et 9.

    sous.ensemble = rbind(G, D)

    x.G <- -median(G[, 1]) #médiane des x à droite.
    x.D <- -median(D[, 1]) #médiane des x à gauche.
    y.G <- -median(G[, 2]) #médiane des y à droite.
    y.D <- -median(D[, 2]) #médiane des y à gauche.

    b.0 <- -((y.D - y.G)/(x.D - x.G)) #premier estimé de la pente.
    ee.0 <- -(y.ordon - (b.0 * x.ordon)) #valeurs résiduelles.
    ee.D.0 <- -median(ee.0[7 : 10]) #médiane des résidus à droite.
    ee.G.0 <- -median(ee.0[1 : 4]) #médiane des résidus à gauche.
    ee.b.0 <- -((ee.D.0 - ee.G.0)/(x.D - x.G)) #pente des résidus.

    bb.1 <- -b.0 + ee.b.0 + correction #deuxième estimé de la pente avec correction si les
      médianes
      #des résidus sont de même signes.

    ee.1 <- -(y.ordon - (bb.1 * x.ordon)) #valeurs résiduelles.
    ee.D.1 <- -median(ee.1[7 : 10]) #médiane des résidus à droite.

```

```

ee.G.1 < -median(ee.1[1 : 4]) #médiane des résidus à gauche.
ee.b.1 < -((ee.D.1 - ee.G.1)/(x.D - x.G)) #pente des résidus.
note < -1} #Indicateur de la position de x.10.
else{if(x[10] == x[4]){
y.ordon = y[order(x)] #Ordonne les observations y selon l'ordre des x.
x.ordon = sort(x) #Ordonne les observations x en ordre croissant.
G = cbind(x.ordon[1 : 5], y.ordon[1 : 5])
      #correspond aux valeurs originales indicées 1,2,3,4 et 10 (pas nécessairement dans
      cet ordre)
C = cbind("vide", "vide") #vide.
D = cbind(x.ordon[6 : 10], y.ordon[6 : 10]) #correspond aux valeurs originales indicées
      5,6,7,8 et 9.
sous.ensemble = rbind(G, D)
x.G < -median(G[, 1]) #médiane des x à droite.
x.D < -median(D[, 1]) #médiane des x à gauche.
y.G < -median(G[, 2]) #médiane des y à droite.
y.D < -median(D[, 2]) #médiane des y à gauche.
b.0 < -((y.D - y.G)/(x.D - x.G)) #premier estimé de la pente.
ee.0 < -(y.ordon - (b.0 * x.ordon)) #valeurs résiduelles.
ee.D.0 < -median(ee.0[6 : 10]) #médiane des résidus à droite.
ee.G.0 < -median(ee.0[1 : 5]) #médiane des résidus à gauche.
ee.b.0 < -((ee.D.0 - ee.G.0)/(x.D - x.G)) #pente des résidus.
bb.1 < -b.0 + ee.b.0 + correction #deuxième estimé de la pente avec correction si les
      médianes
      #des résidus sont de même signes.
ee.1 < -(y.ordon - (bb.1 * x.ordon)) #valeurs résiduelles.

```

```

ee.D.1 < -median(ee.1[6 : 10]) #médiane des résidus à droite.
ee.G.1 < -median(ee.1[1 : 5]) #médiane des résidus à gauche.
ee.b.1 < -((ee.D.1 - ee.G.1)/(x.D - x.G)) #pente des résidus.
note < -2} #Indicateur de la position de x.10.
else{if((x[4] < x[10]&x[10] < x[6])){
y.ordon = y[order(x)] #Ordonne les observations y selon l'ordre des x.
x.ordon = sort(x) #Ordonne les observations x en ordre croissant.
G = cbind(x.ordon[1 : 4], y.ordon[1 : 4])
      #correspond aux valeurs originales indicées 1,2,3 et 4
C = cbind(x.ordon[5 : 6], y.ordon[5 : 6]) #correspond aux valeurs originales indicées 5
      et 10.
D = cbind(x.ordon[7 : 10], y.ordon[7 : 10]) #correspond aux valeurs originales indicées
      6,7,8 et 9.
sous.ensemble = rbind(G, D)
x.G < -median(G[, 1]) #médiane des x à droite.
x.D < -median(D[, 1]) #médiane des x à gauche.
y.G < -median(G[, 2]) #médiane des y à droite.
y.D < -median(D[, 2]) #médiane des y à gauche.
b.0 < -((y.D - y.G)/(x.D - x.G)) #premier estimé de la pente.
ee.0 < -(y.ordon - (b.0 * x.ordon)) #valeurs résiduelles.
ee.D.0 < -median(ee.0[7 : 10]) #médiane des résidus à droite.
ee.G.0 < -median(ee.0[1 : 4]) #médiane des résidus à gauche.
ee.b.0 < -((ee.D.0 - ee.G.0)/(x.D - x.G)) #pente des résidus.
bb.1 < -b.0 + ee.b.0 + correction #deuxième estimé de la pente avec correction si les
      médianes
      #des résidus sont de même signes.

```

```

ee.1 < -(y.ordon - (bb.1 * x.ordon)) #valeurs résiduelles.
ee.D.1 < -median(ee.1[7 : 10]) #médiane des résidus à droite.
ee.G.1 < -median(ee.1[1 : 4]) #médiane des résidus à gauche.
ee.b.1 < -((ee.D.1 - ee.G.1)/(x.D - x.G)) #pente des résidus.
note < -3} #Indicateur de la position de x.10.
else{if(x[10] == x[6]){
y.ordon = y[order(x)] #Ordonne les observations y selon l'ordre des x.
x.ordon = sort(x) #Ordonne les observations x en ordre croissant.
G = cbind(x.ordon[1 : 5], y.ordon[1 : 5])
      #correspond aux valeurs originales indicées 1,2,3,4 et 5
C = cbind("vide", "vide") #vide.
D = cbind(x.ordon[6 : 10], y.ordon[6 : 10]) #correspond aux valeurs originales indicées
      10,6,7,8 et 9
      #(pas nécessairement dans cet ordre).
sous.ensemble = rbind(G, D)
x.G < -median(G[, 1]) #médiane des x à droite.
x.D < -median(D[, 1]) #médiane des x à gauche.
y.G < -median(G[, 2]) #médiane des y à droite.
y.D < -median(D[, 2]) #médiane des y à gauche.
b.0 < -((y.D - y.G)/(x.D - x.G)) #premier estimé de la pente.
ee.0 < -(y.ordon - (b.0 * x.ordon)) #valeurs résiduelles.
ee.D.0 < -median(ee.0[6 : 10]) #médiane des résidus à droite.
ee.G.0 < -median(ee.0[1 : 5]) #médiane des résidus à gauche.
ee.b.0 < -((ee.D.0 - ee.G.0)/(x.D - x.G)) #pente des résidus.
bb.1 < -b.0 + ee.b.0 + correction #deuxième estimé de la pente avec correction si les
      médianes
      #des résidus sont de même signes.

```

```

ee.1 < -(y.ordon - (bb.1 * x.ordon)) #valeurs résiduelles.
ee.D.1 < -median(ee.1[6 : 10]) #médiane des résidus à droite.
ee.G.1 < -median(ee.1[1 : 5]) #médiane des résidus à gauche.
ee.b.1 < -((ee.D.1 - ee.G.1)/(x.D - x.G)) #pente des résidus.
note < -4} #Indicateur de la position de x.10.
else{if(x[10] > x[7]){
y.ordon = y[order(x)] #Ordonne les observations y selon l'ordre des x.
x.ordon = sort(x) #Ordonne les observations x en ordre croissant.
G = cbind(x.ordon[1 : 3], y.ordon[1 : 5])
      #correspond aux valeurs originales indicées 1,2 et 3
C = cbind(x.ordon[4 : 7], y.ordon[4 : 7]) #correspond aux valeurs originales indicées
      4,5,6 et 7.
D = cbind(x.ordon[8 : 10], y.ordon[8 : 10]) #correspond aux valeurs originales indicées
      8,9 et 10
      #(pas nécessairement dans cet ordre).
sous.ensemble = rbind(G, D)
x.G < -median(G[, 1]) #médiane des x à droite.
x.D < -median(D[, 1]) #médiane des x à gauche.
y.G < -median(G[, 2]) #médiane des y à droite.
y.D < -median(D[, 2]) #médiane des y à gauche.
b.0 < -((y.D - y.G)/(x.D - x.G)) #premier estimé de la pente.
ee.0 < -(y.ordon - (b.0 * x.ordon)) #valeurs résiduelles.
ee.D.0 < -median(ee.0[8 : 10]) #médiane des résidus à droite.
ee.G.0 < -median(ee.0[1 : 3]) #médiane des résidus à gauche.
ee.b.0 < -((ee.D.0 - ee.G.0)/(x.D - x.G)) #pente des résidus.

```

```

bb.1 < -b.0 + ee.b.0 + correction #deuxième estimé de la pente avec correction si les
    médianes
    #des résidus sont de même signes.
ee.1 < -(y.ordon - (bb.1 * x.ordon)) #valeurs résiduelles.
ee.D.1 < -median(ee.1[8 : 10]) #médiane des résidus à droite.
ee.G.1 < -median(ee.1[1 : 3]) #médiane des résidus à gauche.
ee.b.1 < -((ee.D.1 - ee.G.1)/(x.D - x.G)) #pente des résidus.
note < -5}}}} #Indicateur de la position de x.10.
colnames(sous.ensemble) = c("x", "y")
resultats < -c(x.D, x.G, b.0, bb.1, ee.b.0, ee.b.1, note)
names(resultats) < -c("med(x)àdroite", "med(x)àgauche", "b.0", "b.1", "penteresid0",
    "penteresid1", "localisationdex.10")
if(choix.resultat == 1){return(resultats)}
else{return(sous.ensemble)}}

```

#La fonction ci-dessous assure que $e(b_0)$ et $e(b_1)$ sont toujours de signes opposés.

```

assignation < -function(b.0, b.1, b.vr.0, b.vr.1){
if(sign(valeur.initiale[5]) + sign(valeur.initiale[6]) == 0){
    # Les médianes des erreurs sont de signes opposés
un = b.0
deux = b.1
trois = b.vr.0
quatre = b.vr.1
resultat = c(un, deux, trois, quatre)}

```



```

else{if(sign(valeur.initiale[5]) + sign(valeur.initiale[6]) == -2){
    #Les médianes des erreurs sont de même signe négatif.
    #Une correction de l'estimé b.1 s'impose.
    valeur.initiale < -Initial.Resis.Line(x = x, y = y, n = length(x),
        correction = -1, choix.resultat = 1)
    un = valeur.initiale[3]
    deux = valeur.initiale[4]
    trois = valeur.initiale[5]
    quatre = valeur.initiale[6]
    resultat = c(un, deux, trois, quatre)}
else{ #Les médianes des erreurs sont de même signe positif.
    #Une correction de l'estimé b.1 s'impose.
    valeur.initiale < -Initial.Resis.Line(x = x, y = y, n = length(x),
        correction = 1, choix.resultat = 1)
    un = valeur.initiale[3]
    deux = valeur.initiale[4]
    trois = valeur.initiale[5]
    quatre = valeur.initiale[6]
    resultat = c(un, deux, trois, quatre)}}
resultat}

```

#La fonction ci-dessous assure que, durant le processus itératif, $e(b_k)^+$ et $e(b_k)^-$ sont toujours de signes opposés.

```

reassignment < -function(pente.resid.k, petit.resid, grand.resid){

```

```

if(sign(pente.resid.k) == sign(petit.resid)){
  c(b.petit.resid < -b.k, petit.resid < -pente.resid.k)}
else{c(b.grand.resid < -b.k, grand.resid < -pente.resid.k)}
c(petit.resid, grand.resid, b.petit.resid, b.grand.resid)}

```

#Le procédé itératif permettant de calculer l'estimé de la pente β .

```

long < -length(x)
valeur.initiale = Initial.Resis.Line(x = x, y = y, n = long, correction = 0, choix.resultat =
  1)
Sous.Ensemble = Initial.Resis.Line(x = x, y = y, n = long, correction = 0, choix.resultat =
  2)
distribution = assignation(b.0 = valeur.initiale[3], b.1 = valeur.initiale[4],
  b.vr.0 = valeur.initiale[5], b.vr.1 = valeur.initiale[6])
b.0 = distribution[1]
b.1 = distribution[2]
b.v.0 = distribution[3]
b.vr.1 = distribution[4]
ifelse(abs(b.vr.0) < abs(b.vr.1), c(b.petit.resid = b.0, b.grand.resid = b.1,
  petit.resid = b.vr.0, grand.resid = b.vr.1), c(b.petit.resid = b.1, b.grand.resid =
  b.0,
  petit.resid = b.vr.1, grand.resid = b.vr.0))
vecteur.pente = numeric()
vecteur.pente = c(vecteur.pente, b.0, b.1)
compteur = 2

```

```

k = 10

ifelse(petit.resid == 0 & grand.resid == 0, vecteur.pente[compteur],
      longueur = length(Sous.Ensemble[, 1]))

while(k > 0.0000001){

b.k = b.petit.resid - petit.resid * ((b.petit.resid - b.grand.resid) / (petit.resid - grand.resid))

vecteur.pente < -c(vecteur.pente, b.k)

residuelles < -(y - (b.k * x))

resid < -(Sous.Ensemble[, 2] - (b.k * Sous.Ensemble[, 1]))

med.resid.r < -median(resid[((longueur/2) + 1) : longueur])

med.resid.l < -median(resid[1 : (longueur/2)])

pente.resid.k < -((med.resid.r - med.resid.l) / (valeur.initiale[1] - valeur.initiale[2]))

redistribution < -reassignation(pente.resid.k, petit.resid, grand.resid)

petit.resid < -redistribution[1]

grand.resid < -redistribution[2]

b.petit.resid < -redistribution[3]

b.grand.resid < -redistribution[4]

compteur < -compteur + 1

k = abs(vecteur.pente[compteur] - vecteur.pente[compteur - 1])

b.DR = vecteur.pente[compteur]

```

#Estimation du paramètre β par la méthode d'estimation bayésienne avec paramètre d'échelle σ connu. Dans un premier temps, on suppose que la variable aléatoire Z_i est distribuée selon une loi $t_4(\beta; \frac{2,2765}{x_i})$, c'est-à-dire une loi de Student avec 4 degrés de liberté, avec paramètre de position β et paramètre d'échelle connu égal à $\frac{2,2765}{x_i}$.

#Définitions des fonctions permettant de calculer les densités *a posteriori*.

```
#Les données sans le point P.10
Y = c(2.4, 0.5, 7.6, 10.3, 6.4, 9.2, 10.7, 15.6, 13.6)
X = c(1.1, 1.2, 1.5, 3, 3.7, 3.9, 3.9, 5, 5.7)

#vraisemblance des erreurs avec paramètre d'échelle sigma fixé
student = function(err){1/((1 + (err^2)/(4))^(5/2))}

#densité a posteriori de beta
pi.S1 = function(b, sigma2){
  z = y/x
  prod(student(err = (z - b)/sigma2 * x)/sigma2 * x)}

integre.S1 = function(sigma, db = .1, min.b = 0, max.b = 50){
  b = seq(min.b + db/2, max.b - db/2, by = db)
  f2 = apply(as.matrix(b), 1, pi.S1, sigma2 = sigma)
  cten = sum(f2) * db
  f2 = f2/cten
  return(f2)}

#densité a posteriori de beta sans le point P10
pi.S2 = function(b, sigma2){
  Z = Y/X
  prod(student(err = (Z - b)/sigma2 * X)/sigma2 * X)}
```

```

integre.S2 = function(sigma, db = .1, min.b = 0, max.b = 50){
  b = seq(min.b + db/2, max.b - db/2, by = db)
  f2 = apply(as.matrix(b), 1, pi.S2, sigma2 = sigma)
  cten = sum(f2) * db
  f2 = f2/cten
  return(f2)}

```

#densité a posteriori de beta pour le i-ème point

```

pi.S3 = function(b, sigma2, i){
  z = y[i]/x[i]
  student(err = (z - b)/sigma2 * x[i])/sigma2 * x[i]}

```

```

integre.S3 = function(sigma, j, db = .1, min.b = 0, max.b = 50){
  b = seq(min.b + db/2, max.b - db/2, by = db)
  f2 = apply(as.matrix(b), 1, pi.S3, sigma2 = sigma, i = j)
  cten = sum(f2) * db
  f2 = f2/cten
  return(f2)}

```

```

valeur.sigma = 2.2765

```

#Définition du domaine d'intégration

```

db = .001
min.b = 0
max.b = 50
b = seq(min.b + db/2, max.b - db/2, by = db)

```

```
#Estimé bayésien de la pente donné par l'espérance a posteriori de  $\beta$ .
b.student = sum(b * integre.S1(sigma = valeur.sigma, db = .001, min.b = 0, max.b =
50)) * db
```

#Estimation du paramètre β par la méthode d'estimation bayésienne avec paramètre d'échelle σ connu. On suppose maintenant que la variable aléatoire Z_i est distribuée selon une loi F_{GEP} , c'est-à-dire une loi GEP de type V avec paramètre de position β et paramètre d'échelle connu $\frac{4,2773}{x_i}$.

```
#Définitions des fonctions permettant de calculer les densités a posteriori.
```

```
#Les données sans le point P.10
Y = c(2.4, 0.5, 7.6, 10.3, 6.4, 9.2, 10.7, 15.6, 13.6)
X = c(1.1, 1.2, 1.5, 3, 3.7, 3.9, 3.9, 5, 5.7)

#vraisemblance des erreurs avec paramètre d'échelle sigma fixé
GEP = fonction(err){(abs(err) + 2)^(-1) * (log(abs(err) + 2))^(-4)}

#densité a posteriori de beta
pi.GEP1 = fonction(b, sigma2){
  z = y/x
  prod(GEP(err = (z - b)/sigma2 * x)/sigma2 * x)}
```

```

integre.GEP1 = function(sigma, db = .1, min.b = 0, max.b = 50){
  b = seq(min.b + db/2, max.b - db/2, by = db)
  f2 = apply(as.matrix(b), 1, pi.GEP1, sigma2 = sigma)
  cten = sum(f2) * db
  f2 = f2/cten
  return(f2)}

```

#densité a posteriori de beta sans le point P10

```

pi.GEP2 = function(b, sigma2){
  Z = Y/X
  prod(GEP(err = (Z - b)/sigma2 * X)/sigma2 * X)}

```

```

integre.GEP2 = function(sigma, db = .1, min.b = 0, max.b = 50){
  b = seq(min.b + db/2, max.b - db/2, by = db)
  f2 = apply(as.matrix(b), 1, pi.GEP2, sigma2 = sigma)
  cten = sum(f2) * db
  f2 = f2/cten
  return(f2)}

```

#densité a posteriori de beta pour le i-ème point

```

pi.GEP3 = function(b, sigma2, i){
  z = y[i]/x[i]
  GEP(err = (z - b)/sigma2 * x[i])/sigma2 * x[i]}

```

```

integre.GEP3 = function(sigma, j, db = .001, min.b = 0, max.b = 50){
  b = seq(min.b + db/2, max.b - db/2, by = db)
  f2 = apply(as.matrix(b), 1, pi.GEP3, sigma2 = sigma, i = j)

```

```
cten = sum(f2) * db
```

```
f2 = f2/cten
```

```
return(f2)}
```

```
valeur.sigma = 4.2773
```

```
#Définition du domaine d'intégration
```

```
db = .001
```

```
min.b = 0
```

```
max.b = 50
```

```
b = seq(min.b + db/2, max.b - db/2, by = db)
```

```
#Estimé bayésien de la pente donné par l'espérance a posteriori de  $\beta$ .
```

```
b.GEP = sum(b*integre.GEP1(sigma = valeur.sigma, db = .001, min.b = 0, max.b =  
50)) * db
```

```
#Estimation des paramètres  $\beta$  et  $\sigma$  par la méthode d'estimation bayésienne simultanée.
```

On suppose que la variable aléatoire Z_i est distribuée selon une loi $t_4 \left(\beta; \sigma \left(\frac{0.9106}{x_i} \right) \right)$, c'est-à-dire une loi de student avec 4 degrés de liberté, avec paramètre de position β et paramètre d'échelle inconnu $\sigma \left(\frac{0.9106}{x_i} \right)$.

```
#vraisemblance des erreurs
```

```
student = function(err){1/((1 + (err^2)/(4))^(5/2))}
```

```
#Calcul des densités a posteriori de beta et de sigma, respectivement
```

```
pi.S1 = function(b, sigma2, tau = 0.9106){
```



```

z = y/x

prod(student(err = (z-b)/(sigma2*tau)*x)/(sigma2*tau)*x)*(1/(sigma2*tau))}

integre.lb = function(sigma, db = .1, min.b = 0, max.b = 50){
  b = seq(min.b + db/2, max.b - db/2, by = db)
  f2 = apply(as.matrix(b), 1, pi.S1, sigma2 = sigma, tau = 0.9106)
  return(f2)}

integre.ls = function(b2, dsigma = .1, min.sigma = 0, max.sigma = 50){
  sigma = seq(min.sigma + dsigma/2, max.sigma - dsigma/2, by = dsigma)
  f2 = apply(as.matrix(sigma), 1, pi.S1, b = b2, tau = 0.9106)
  return(f2)}

tau = 0.9106

#Définition du domaine d'intégration pour l'estimé b
db = .001
min.b = 0
max.b = 50
b = seq(min.b + db/2, max.b - db/2, by = db)

#Définition du domaine d'intégration pour l'estimé σ
dsigma = .001
min.sigma = 0
max.sigma = 50
sigma = seq(min.sigma + dsigma/2, max.sigma - dsigma/2, by = dsigma)

```

```

#Calcul de la constante de normalisation
post.conjoint = 0
for(i in range.sigma){
  print(i)
  dist.posterior = (integrate.lb(sigma = i, db = .001, min.b = 0, max.b = 50))
  post.conjoint = (post.conjoint) + (dist.posterior)}
constante = sum(post.conjoint) * db * dsigma

#Calcul de la densité marginale a posteriori de  $\beta$ 
dist.posterior = sum(integrate.ls(b2 = i, dsigma = .001, min.sigma = 0, max.sigma =
  50)) * dsigma
marg.beta = c(marg.beta, dist.posterior)
post.marginal.beta = (marg.beta)/constante

#Estimé bayésien de la pente donné par l'espérance a posteriori de  $\beta$ .
b.student = (sum(b * post.marginal.beta) * db)

#Calcul de la densité marginale a posteriori de  $\sigma$ 
dist.posterior = sum(integrate.lb(sigma = i, dsigma = .001, min.sigma = 0, max.sigma =
  50)) * db
marg.sigma = c(marg.sigma, dist.posterior)
post.marginal.sigma = (marg.sigma)/constante

#Estimé bayésien du paramètre d'échelle donné par l'espérance a posteriori de  $\sigma$ .
b.student = (sum(b * post.marginal.beta) * db)

#Estimé bayésien de la pente donné par l'espérance a posteriori de  $\beta$ .

```

```
sigma.student = (sum(sigma * post.marginal.sigma) * dsigma)
```

#Estimation des paramètres β et σ par la méthode d'estimation bayésienne simultanée. On suppose que la variable aléatoire Z_i est distribuée selon une loi H_{GEP} , c'est-à-dire une loi GEP de type V avec paramètre de position β et paramètre d'échelle inconnu $\sigma \left(\frac{1.7109}{x_i} \right)$.

```
#vraisemblance des erreurs
```

```
GEP = function(err){(abs(err) + 2)^(-1) * (log(abs(err) + 2))^( -4)}
```

```
#Calcul des densités a posteriori de beta et de sigma, respectivement
```

```
pi.GEP1 = function(b, sigma2, tau = 1.7109){
```

```
z = y/x
```

```
prod(GEP(err = (z - b)/(sigma2 * tau) * x)/(sigma2 * tau) * x * (1/(sigma2 * tau)))}
```

```
integre.1b = function(sigma, db = .1, min.b = 0, max.b = 50){
```

```
b = seq(min.b + db/2, max.b - db/2, by = db)
```

```
f2 = apply(as.matrix(b), 1, pi.GEP1, sigma2 = sigma, tau = 1.7109)
```

```
return(f2)}
```

```
integre.1s = function(b2, dsigma = .1, min.sigma = 0, max.sigma = 50){
```

```
sigma = seq(min.sigma + dsigma/2, max.sigma - dsigma/2, by = dsigma)
```

```
f2 = apply(as.matrix(sigma), 1, pi.GEP1, b = b2, tau = 1.7109)
```

```
return(f2)}
```

```
tau = 1.7109
```

```

#Définition du domaine d'intégration pour l'estimé  $b$ 

db = .001

min.b = 0

max.b = 50

b = seq(min.b + db/2, max.b - db/2, by = db)

#Définition du domaine d'intégration pour l'estimé  $\sigma$ 

dsigma = .001

min.sigma = 0

max.sigma = 50

sigma = seq(min.sigma + dsigma/2, max.sigma - dsigma/2, by = dsigma)


#Calcul de la constante de normalisation

post.conjoint = 0

for(i in range.sigma){

  print(i)

  dist.posterior = (integrate.lb(sigma = i, db = .001, min.b = 0, max.b = 50))

  post.conjoint = (post.conjoint) + (dist.posterior)}

constante = sum(post.conjoint) * db * dsigma


#Calcul de la densité marginale a posteriori de  $\beta$ 

dist.posterior = sum(integrate.lsb(b2 = i, dsigma = .001, min.sigma = 0, max.sigma =
  50)) * dsigma

marg.beta = c(marg.beta, dist.posterior)

post.marginal.beta = (marg.beta)/constante

```

#Estimé bayésien de la pente donné par l'espérance *a posteriori* de β .

$b.GEP = (sum(b * post.marginal.beta) * db)$

#Calcul de la densité marginale *a posteriori* de σ

$dist.posterior = sum(integre.lb(sigma = i, dsigma = .001, min.sigma = 0, max.sigma = 50)) * db$

$marg.sigma = c(marg.sigma, dist.posterior)$

$post.marginal.sigma = (marg.sigma) / constante$

#Estimé bayésien du paramètre d'échelle donné par l'espérance *a posteriori* de σ .

$b.GEP = (sum(b * post.marginal.beta) * db)$

#Estimé bayésien de la pente donné par l'espérance *a posteriori* de β .

$sigma.GEP = (sum(sigma * post.marginal.sigma) * dsigma)$

RÉFÉRENCES

- Andrews, D., P. Bickel, F. Hampel, P. Huber, W. Rogers, et J. Tukey. 1972. *Robust estimates of location*. New Jersey : Princeton University Press.
- Angers, J.-F. 2000. « P-credence and outliers », *Metron*, vol. 58, no. 3-4, p. 81-108.
- Bartlett, M. 1949. « Fitting a straight line when both variables are subject to error », *Biometrics*, vol. 5, no. 3, p. 207-212.
- Birkes, D. et Y. Dodge. 1993. *Alternative methods of regression*. New York : John Wiley and Sons.
- Dawid, P. 1973. « Posterior expectations for large observations », *Biometrika*, vol. 60, no. 3, p. 664-667.
- Desgagné, A. 2005. « Quelques utilisations de la densité GEP en analyse bayésienne sur les familles de position-échelle ». Thèse de Doctorat, Montréal, Université de Montréal.
- Desgagné, A. et J.-F. Angers. 2003. Computational aspect of the generalized exponential power density. Rapport, Centre de recherches mathématiques de l'Université de Montréal, Montréal.
- . 2005. « Importance sampling with the generalized exponential power density », *Statistics and computing*, vol. 15, no. 3, p. 189-196.
- . 2007. « Conflicting information and location parameter inference », *Metron*, vol. 65, no. 1, p. 67-97.
- Dutter, R. 1977. « Algorithms for the Huber estimator in multiple regression », *Computing*, vol. 18, no. 2, p. 167-176.
- Hampel, F. R., E. M. Rochetti, P. J. Rousseeuw, et W. A. Stahel. 1986. *Robust statistics. The approach based on influence functions*. New York : John Wiley and Sons.
- Hill, B. 1974. *On coherence, inadmissibility and inference about many parameters in the theory of least squares*. Coll. « Studies in Bayesian Econometrics and Statistics (In Honor of Leonard J. Savage) », p. 555-584. Amsterdam : North-Holland Publishing Co.
- Hoaglin, D. C., F. Mosteller, et J. W. Tukey. 1983. *Understanding robust and exploratory data analysis*. New York : John Wiley and Sons.

- . 1985. *Exploring data tables, trends and shapes*. New York : John Wiley and Sons.
- Huber, P. J. 1964. « Robust estimation of a location parameter », *The Annals of Mathematical Statistics*, vol. 35, no. 1, p. 73–101.
- . 1981. *Robust Statistics*. New York : John Wiley and Sons.
- Johnstone, I. M. et P. F. Velleman. 1985. « The resistant line and related regression methods », *Journal of the American Statistical Association*, vol. 80, no. 392, p. 1041–1054.
- Kildea, D. G. 1981. « Brown-Mood type median estimators for simple regression models », *The Annals of Statistics*, vol. 9, no. 2, p. 438–442.
- Lange, K. L., R. J. A. Little, et J. M. G. Taylor. 1989. « Robust statistical modeling using the t distribution », *Journal of the American Statistical Association*, vol. 84, no. 408, p. 881–896.
- Lawrence, K. D. et J. L. Arthur. 1990. *Robust statistics. The approach based on influence functions*. New York : Marcel Dekker.
- Lindley, D. V. 1968. « The choice of variables in multiple regression », *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no. 1, p. 31–66.
- Mood, A. M. 1950. *Introduction to the theory of statistics*. New York : McGraw-Hill.
- Nair, K. et M. Shrivastava. 1942. « On a simple method of curve fitting », *Sankhyā*, vol. 11, no. 3, p. 284–300.
- O'Hagan, A. 1979. « On outlier rejection phenomena in bayes inference », *Journal of the Royal*, vol. 41, no. 3, p. 358–367.
- . 1990. « Outliers and credence for location parameter inference », *Journal of the American Statistical Association*, vol. 85, no. 409, p. 172–176.
- Rousseeuw, P. 1984. « Least median of squares regression », *Journal of the American Statistical Association*, vol. 79, no. 388, p. 871–880.
- Rousseeuw, P. et M. Hubert. 1999. « Regression depth », *Journal of the American Statistical Association*, vol. 94, no. 446, p. 388–402.
- Rousseeuw, P. J. et A. M. Leroy. 2003. *Robust regression and outlier detection*. New York : John Wiley and Sons.
- Velleman, P. F. et D. C. Hoaglin. 1981. *Applications, basics and computing of exploratory data analysis*. Boston : Duxbury Press.
- Wald, A. 1940. « Fitting a straight line if both variables are subject to error », *The annals of mathematical statistics*, vol. 11, no. 3, p. 284–300.